

## TWO SAMPLE TESTS FOR HIGH-DIMENSIONAL COVARIANCE MATRICES

BY JUN LI AND SONG XI CHEN<sup>1</sup>

*Iowa State University, and Peking University and Iowa State University*

We propose two tests for the equality of covariance matrices between two high-dimensional populations. One test is on the whole variance–covariance matrices, and the other is on off-diagonal submatrices, which define the covariance between two nonoverlapping segments of the high-dimensional random vectors. The tests are applicable (i) when the data dimension is much larger than the sample sizes, namely the “large  $p$ , small  $n$ ” situations and (ii) without assuming parametric distributions for the two populations. These two aspects surpass the capability of the conventional likelihood ratio test. The proposed tests can be used to test on covariances associated with gene ontology terms.

**1. Introduction.** Modern statistical data are increasingly high dimensional, but with relatively small sample sizes. Genetic data typically carry thousands of dimensions for measurements on the genome. However, due to limited resources available to replicate study objects, the sample sizes are usually much smaller than the dimension. This is the so-called “large  $p$ , small  $n$ ” paradigm. An enduring interest in Statistics is to know if two populations share the same distribution or certain key distributional characteristics, for instance the mean or covariance. The two populations here can refer to two “treatments” in a study. As testing for equality of high-dimensional distributions is far more challenging than that for the fixed-dimensional data, testing for equality of key characteristics of the distributions is more achievable and desirable due to easy interpretation. There has been a set of research on inference for means of high-dimensional distributions either in the context of multiple testing, as in van der Laan and Bryan (2001), Donoho and Jin (2004), Fan, Hall and Yao (2007) and Hall and Jin (2008), or in

---

Received July 2011; revised December 2011.

<sup>1</sup>Supported by a NSFC Key Grant 111 31002.

*AMS 2000 subject classifications.* Primary 62H15; secondary 62G10, 62G20.

*Key words and phrases.* High-dimensional covariance, large  $p$  small  $n$ , likelihood ratio test, testing for gene-sets.

This is an electronic reprint of the original article published by the Institute of Mathematical Statistics in *The Annals of Statistics*, 2012, Vol. 40, No. 2, 908–940. This reprint differs from the original in pagination and typographic detail.

the context of simultaneous multivariate testing as in Bai and Saranadasa (1996) and Chen and Qin (2010). See also Huang, Wang and Zhang (2005), Fan, Peng and Huang (2005) and Zhang and Huang (2008) for inference on high-dimensional conditional means.

In addition to detecting difference among the population means, there is a strong motivation for comparing dependence among components of random vectors under different treatments, as high data dimensions can potentially increase the complexity of the dependence. In genomic studies, genetic measurements, either the micro-array expressions or the single nucleotide polymorphism (SNP) counts, may have an internal structure dictated by the genetic networks of living cells. And the variations and dependence among the measurements of the genes may be different under different biological conditions and treatments. For instance, some genes may be tightly correlated in the normal or less severe conditions, but they can become decoupled due to certain disease progression; see Shedden and Taylor (2004) for a discussion.

There have been advances on inference for high-dimensional covariance matrices. The probability limits and the limiting distributions of extreme eigenvalues of the sample covariance matrix based on the random matrix theory are developed in Bai (1993), Bai and Yin (1993), Tracy and Widom (1996), Johnstone (2001) and El Karoui (2007), Johnstone and Lu (2009), Bai and Silverman (2010) and others. Wu and Pourahmadi (2003) and Bickel and Levina (2008a, 2008b) proposed consistent estimators to the population covariance matrices by either truncation or Cholesky decomposition. Fan, Fan and Lv (2008), Lam and Yao (2011) and Lam, Yao and Bathia (2011) considered covariance estimation under factor models. There are also developments in conducting LASSO-type regularization estimation of high-dimensional covariances in Huang et al. (2006) and Rothman, Levina and Zhu (2010). Despite these developments, it is still challenging to transform these results to test procedures on high-dimensional covariance matrices.

As part of the effort in discovering significant differences between two high-dimensional distributions, we develop in this paper two-sample test procedures on high-dimensional covariance matrices. Let  $X_{i1}, \dots, X_{in_i}$  be an independent and identically distributed sample drawn from a  $p$ -dimensional distribution  $F_i$ , for  $i = 1$  and  $2$ , respectively. Here the dimensionality  $p$  can be a lot larger than the two sample sizes  $n_1$  and  $n_2$  so that  $p/n_i \rightarrow \infty$ . Let  $\mu_i$  and  $\Sigma_i$  be, respectively, the mean vector and variance-covariance matrix of the  $i$ th population. The primary interest is to test

$$(1.1) \quad H_{0a}: \Sigma_1 = \Sigma_2 \quad \text{versus} \quad H_{1a}: \Sigma_1 \neq \Sigma_2.$$

Testing for the above high-dimensional hypotheses is a nontrivial statistical problem. Designed for fixed-dimensional data, the conventional likelihood ratio test [see Anderson (2003) for details] may be used for the above hy-

pothesis under  $p \leq \min\{n_1, n_2\}$ . If we let

$$\bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij} \quad \text{and} \quad Q_i = \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)',$$

then the likelihood ratio (LR) statistic for  $H_{0a}$  is

$$\lambda_n = \frac{\prod_{i=1}^2 |Q_i|^{(1/2)n_i}}{|Q|^{(1/2)n}} \frac{n^{(1/2)pn}}{\prod_{i=1}^2 n_i^{(1/2)pn_i}},$$

where  $Q = Q_1 + Q_2$  and  $n = n_1 + n_2$ . However, when  $p > \min\{n_1, n_2\}$ , at least one of the sample covariance matrices  $Q_i/(n_i - 1)$  is singular [Dykstra (1970)]. This causes the LR statistic  $-2\log(\lambda_n)$  to be either infinite or undefined, which fundamentally alters the limiting behavior of the LR statistic. In an important development, Bai et al. (2009) demonstrated that even when  $p \leq \min\{n_1, n_2\}$  where  $\lambda_n$  is properly defined, the test encounters a power loss if  $p \rightarrow \infty$  in such a manner that  $p/n_i \rightarrow c_i \in (0, 1)$  for  $i = 1$  and 2. By employing the theory of large dimensional random matrices, Bai et al. (2009) proposed a correction to the LR statistic and demonstrated that the corrected test is valid under  $p/n_i \rightarrow c_i \in (0, 1)$ . Schott (2007) proposed a test based on a metric that measures the difference between the two sample covariance matrices by assuming  $p/n_i \rightarrow c_i \in [0, \infty)$  and the normal distributions. There are also one sample tests for a high-dimensional variance-covariance  $\Sigma$ . Ledoit and Wolf (2002) and Chen, Zhang and Zhong (2010) introduced tests for  $\Sigma$  being sphericity and identity for normally distributed random vectors. Ledoit and Wolf (2004) considered a class of covariance estimators which are convex sums of  $S_n$  and  $I_p$  under moderate dimensionality ( $p/n \rightarrow c$ ). Cai and Jiang (2011) developed tests for  $\Sigma$  having a banded diagonal structure based on random matrix theory. Lan et al. (2010) developed a bias-corrected test to examine the significance of the off-diagonal elements of the residual covariance matrix. All these tests assume either normality or moderate dimensionality such that  $p/n \rightarrow c$  for a finite constant  $c$ , or both.

We develop in this paper two-sample tests on high-dimensional variance-covariances without the normality assumption while allowing the dimension to be much larger than the sample sizes. In addition to testing for the whole variance-covariance matrices, we propose a test on the equality of off-diagonal sub-matrices in  $\Sigma_1$  and  $\Sigma_2$ . The interest on such a test arises naturally in applications, when we are interested in knowing if two segments of the high-dimensional data share the same covariance between the two treatments. We will argue in Section 3 that the two tests on the whole covariance and the off-diagonal sub-matrices may be used collectively to reduce the dimensionality of the testing problem.

This paper is organized as follows. We propose the two-sample test for the whole covariance matrices in Section 2 which includes the asymptotic

normality of the test statistic and a power evaluation. Properties of the test for the off-diagonal sub-matrices are reported in Section 3. Results from simulation studies are outlined in Section 4. Section 5 demonstrates how to apply the proposed tests on a gene ontology data set for acute lymphoblastic leukemia. All technical details are relegated to Section 6.

**2. Test for high-dimensional variance-covariance.** The test statistic for the hypothesis (1.1) is formulated by targeting on  $\text{tr}\{(\Sigma_1 - \Sigma_2)^2\}$ , the squared Frobenius norm of  $\Sigma_1 - \Sigma_2$ . Although the Frobenius norm is large in magnitude compared with other matrix norms, using it for testing brings two advantages. One is that test statistics based on the norm are relatively easier to be analyzed than those based on the other norm, which is especially the case when considering the limiting distribution of the test statistics. The latter renders formulations of test procedures and power analysis, as we will demonstrate later. The other advantage is that it can be used to directly target on certain sections of the covariance matrix as shown in the next section. The latter would be hard to accomplish with other norms.

As  $\text{tr}\{(\Sigma_1 - \Sigma_2)^2\} = \text{tr}(\Sigma_1^2) + \text{tr}(\Sigma_2^2) - 2\text{tr}(\Sigma_1\Sigma_2)$ , we will construct estimators for each term. It is noted that  $\text{tr}(S_{nh}^2)$ , where  $S_{nh}$  is the sample covariance of the  $h$ th sample, is a poor estimator of  $\text{tr}(\Sigma_h^2)$  under high dimensionality. The idea is to streamline terms in  $\text{tr}(S_{nh}^2)$  so as to make it unbiased to  $\text{tr}(\Sigma_h^2)$  and easier to analyze in subsequent asymptotic evaluations. We consider U-statistics of form  $\frac{1}{n_h(n_h-1)} \sum_{i \neq j} (X'_{hi}X_{hj})^2$  which is unbiased if  $\mu_h = 0$ . To account for  $\mu_h \neq 0$ , we subtract two other U-statistics of order three and four, respectively, using an approach dated back to Glasser (1961, 1962). Specifically, we propose

$$\begin{aligned} A_{n_h} = & \frac{1}{n_h(n_h-1)} \sum_{i \neq j} (X'_{hi}X_{hj})^2 - \frac{2}{n_h(n_h-1)(n_h-2)} \sum_{i,j,k}^* X'_{hi}X_{hj}X'_{hj}X_{hk} \\ (2.1) \quad & + \frac{1}{n_h(n_h-1)(n_h-2)(n_h-3)} \sum_{i,j,k,l}^* X'_{hi}X_{hj}X'_{hk}X_{hl} \end{aligned}$$

to estimate  $\text{tr}(\Sigma_h^2)$ . Throughout this paper we use  $\sum^*$  to denote summation over mutually distinct indices. For example,  $\sum_{i,j,k}^*$  means summation over  $\{(i,j,k) : i \neq j, j \neq k, k \neq i\}$ . Similarly, the estimator for  $\text{tr}(\Sigma_1\Sigma_2)$  is

$$\begin{aligned} C_{n_1n_2} = & \frac{1}{n_1n_2} \sum_i \sum_j (X'_{1i}X_{2j})^2 - \frac{1}{n_1n_2(n_1-1)} \sum_{i,k}^* \sum_j X'_{1i}X_{2j}X'_{2j}X_{1k} \\ (2.2) \quad & - \frac{1}{n_1n_2(n_2-1)} \sum_{i,k}^* \sum_j X'_{2i}X_{1j}X'_{1j}X_{2k} \end{aligned}$$

$$+ \frac{1}{n_1 n_2 (n_1 - 1)(n_2 - 1)} \sum_{i,k}^* \sum_{j,l}^* X'_{1i} X_{2j} X'_{1k} X_{2l}.$$

There are other ways to attain estimators for  $\text{tr}(\Sigma_h^2)$  and  $\text{tr}(\Sigma_1 \Sigma_2)$ . In fact, there is a family of estimators for  $\text{tr}(\Sigma_h^2)$  in the form of  $\text{tr}(S_h^2) - \alpha_{n_h} \sum_{i=1}^{n_h} \text{tr}\{(X_{hi} X'_{hi} - S_h)^2\}$  where  $\alpha_{n_h} = \alpha/n_h^2$  for any constant  $\alpha$ . A family can be similarly formulated for  $\text{tr}(\Sigma_1 \Sigma_2)$ . It can be shown that this family of estimators is asymptotically equivalent to the proposed  $A_{n_h}$  in the sense that they share the same leading order term. However, this family is more complex than the proposed.

The test statistic is

$$(2.3) \quad T_{n_1, n_2} = A_{n_1} + A_{n_2} - 2C_{n_1 n_2}$$

which is unbiased for  $\text{tr}\{(\Sigma_1 - \Sigma_2)^2\}$ . Besides the unbiasedness,  $T_{n_1, n_2}$  is invariant under the location shift and orthogonal rotation. This means that we can assume without loss of generality that  $E(X_{ij}) = 0$  in the rest of the paper. As noted by a reviewer, the computation of  $T_{n_1, n_2}$  would be extremely heavy if the sample sizes  $n_h$  are very large. Indeed, the computation burden comes from the last two sums in  $A_{n_h}$  and the last three in  $C_{n_1, n_2}$ , where the numbers of terms in the summations are in the order of  $n_h^3$  or  $n_h^4$ , respectively. Although the main motivation was the “large  $p$  small  $n$ ” situations, we nevertheless require  $n_h \rightarrow \infty$  in our asymptotic justifications. A solution to alleviate the computation burden can be found by noting that the last two terms in  $A_{n_h}$  and the last three in  $C_{n_1, n_2}$  are all of smaller order than the first, under the assumption of  $\mu_h = 0$ . This means that we can first transform each datum  $X_{hi}$  to  $X_{hi} - \bar{X}_{n_h}$ , and then compute only the first term in (2.1) and (2.2). These will reduce the computation to  $O(n_h^2)$  without affecting the asymptotic normality. The only price paid for such an operation is that the modified statistic is no longer unbiased.

To establish the limiting distribution of  $T_{n_1, n_2}$  so as to establish the two sample test for the variance-covariance, we assume the following conditions:

- A1. As  $\min\{n_1, n_2\} \rightarrow \infty$ ,  $n_1/(n_1 + n_2) \rightarrow \rho$  for a fixed constant  $\rho \in (0, 1)$ .
- A2. As  $\min\{n_1, n_2\} \rightarrow \infty$ ,  $p = p(n_1, n_2) \rightarrow \infty$ , and for any  $k$  and  $l \in \{1, 2\}$ ,  $\text{tr}(\Sigma_k \Sigma_l) \rightarrow \infty$  and

$$(2.4) \quad \text{tr}\{(\Sigma_i \Sigma_j)(\Sigma_k \Sigma_l)\} = o\{\text{tr}(\Sigma_i \Sigma_j) \text{tr}(\Sigma_k \Sigma_l)\}.$$

- A3. For each  $i = 1$  or  $2$ ,  $X_{ij} = \Gamma_i Z_{ij} + \mu_i$  where  $\Gamma_i$  is a  $p \times m_i$  matrix such that  $\Gamma_i \Gamma_i' = \Sigma_i$ ,  $\{Z_{ij}\}_{j=1}^{n_i}$  are independent and identically distributed (i.i.d.)  $m_i$ -dimensional random vectors with  $m_i \geq p$  and satisfy  $E(Z_{ij}) = 0$ ,  $\text{Var}(Z_{ij}) = I_{m_i}$ , the  $m_i \times m_i$  identity matrix. Furthermore, if write  $Z_{ij} = (z_{ij1}, \dots, z_{ijm_i})'$ , then each  $z_{ijk}$  has finite 8th moment,  $E(z_{ijk}^4) = 3 + \Delta_i$  for some constant  $\Delta_i$  and for any positive integers  $q$  and  $\alpha_l$  such that  $\sum_{l=1}^q \alpha_l \leq 8E(z_{ijl_1}^{\alpha_1} \cdots z_{ijl_q}^{\alpha_q}) = E(z_{ijl_1}^{\alpha_1}) \cdots E(z_{ijl_q}^{\alpha_q})$  for any  $l_1 \neq l_2 \neq \cdots \neq l_q$ .

While Condition A1 is of standard for two-sample asymptotic analysis, A2 spells the extent of high dimensionality and the dependence which can be accommodated by the proposed tests. A key aspect is that it does not impose any explicit relationships between  $p$  and the sample sizes, but rather requires a quite mild (2.4) regarding the covariances. To appreciate (2.4), we note that if  $i = j = k = l$ , it has the form of  $\text{tr}(\Sigma_i^4) = o\{\text{tr}^2(\Sigma_i^2)\}$ , which is valid if all the eigenvalues of  $\Sigma_i$  are uniformly bounded. Condition (2.4) also makes the asymptotic study of the test statistic manageable under high dimensionality. We note here that requiring  $\text{tr}(\Sigma_k \Sigma_l) \rightarrow \infty$  is a precursor to (2.4). We do not assume specific parametric distributions for the two samples. Instead, a general multivariate model is assumed in A3 which was advocated in Bai and Saranadasa (1996) for testing high dimensional means. The model resembles that of the factor model with  $Z_i$  representing the factors, except that here we allow the number of factor  $m_i$  at least as large as  $p$ . This provides flexibility in accommodating a wider range of multivariate distributions for the observed data  $X_{ij}$ .

Derivations leading to (6.4) in Section 6 show that, under A2 and A3, the leading order variance of  $T_{n_1, n_2}$  under either  $H_{0a}$  or  $H_{1a}$  is

$$\begin{aligned}
 \sigma_{n_1, n_2}^2 = & \sum_{i=1}^2 \left[ \frac{4}{n_i^2} \text{tr}^2(\Sigma_i^2) + \frac{8}{n_i} \text{tr}\{(\Sigma_i^2 - \Sigma_1 \Sigma_2)^2\} \right. \\
 (2.5) \quad & \left. + \frac{4\Delta_i}{n_i} \text{tr}\{\Gamma'_i(\Sigma_1 - \Sigma_2)\Gamma_i \circ \Gamma'_i(\Sigma_1 - \Sigma_2)\Gamma_i\} \right] \\
 & + \frac{8}{n_1 n_2} \text{tr}^2(\Sigma_1 \Sigma_2),
 \end{aligned}$$

where  $A \circ B = (a_{ij}b_{ij})$  for two matrices  $A = (a_{ij})$  and  $B = (b_{ij})$ . Note that for any symmetric matrix  $A$ ,  $\text{tr}(A \circ A) \leq \text{tr}(A^2)$ . Hence,

$$\begin{aligned}
 \text{tr}\{\Gamma'_1(\Sigma_1 - \Sigma_2)\Gamma_1 \circ \Gamma'_1(\Sigma_1 - \Sigma_2)\Gamma_1\} & \leq \text{tr}\{(\Sigma_1^2 - \Sigma_1 \Sigma_2)^2\} \quad \text{and} \\
 \text{tr}\{\Gamma'_2(\Sigma_1 - \Sigma_2)\Gamma_2 \circ \Gamma'_2(\Sigma_1 - \Sigma_2)\Gamma_2\} & \leq \text{tr}\{(\Sigma_2^2 - \Sigma_2 \Sigma_1)^2\}.
 \end{aligned}$$

These together with the fact that  $\Delta_i \geq -2$  ensure that  $\sigma_{n_1, n_2}^2 > 0$ . We note that the  $\Gamma_i$ - $Z_{ij}$  pair in Model A3 is not unique, and there are other pairs, say  $\tilde{\Gamma}_i$  and  $\tilde{Z}_{ij}$ , such that  $X_{ij} = \tilde{\Gamma}_i \tilde{Z}_{ij}$ . However, it can be shown that the value of  $\frac{4\Delta_i}{n_i} \text{tr}\{\Gamma'_i(\Sigma_1 - \Sigma_2)\Gamma_i \circ \Gamma'_i(\Sigma_1 - \Sigma_2)\Gamma_i\}$  remains the same.

The following theorem establishes the asymptotic normality of  $T_{n_1, n_2}$ .

**THEOREM 1.** *Under Conditions A1–A3, as  $\min\{n_1, n_2\} \rightarrow \infty$*

$$\sigma_{n_1, n_2}^{-1} [T_{n_1, n_2} - \text{tr}\{(\Sigma_1 - \Sigma_2)^2\}] \xrightarrow{d} N(0, 1).$$

It is noted that under  $H_{0a} : \Sigma_1 = \Sigma_2 = \Sigma$ , say,  $\sigma_{n_1, n_2}^2$  becomes

$$\sigma_{0, n_1, n_2}^2 = 4 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)^2 \text{tr}^2(\Sigma^2).$$

To formulate a test procedure, we need to estimate  $\sigma_{0, n_1, n_2}^2$ . As  $A_{n_1}$  and  $A_{n_2}$  are unbiased estimators of  $\text{tr}(\Sigma_1^2)$  and  $\text{tr}(\Sigma_2^2)$ , respectively, we will use  $\hat{\sigma}_{0, n_1, n_2}^2 = \frac{2}{n_2} A_{n_1} + \frac{2}{n_1} A_{n_2}$  as the estimator. The following theorem shows that  $\hat{\sigma}_{0, n_1, n_2}^2$  is ratio-consistent to  $\sigma_{0, n_1, n_2}^2$ .

**THEOREM 2.** *Under Conditions A1–A3 and  $H_{0a}$ , as  $\min\{n_1, n_2\} \rightarrow \infty$ ,*

$$(2.6) \quad \frac{A_{n_i}}{\text{tr}(\Sigma_i^2)} \xrightarrow{p} 1 \quad \text{for } i = 1 \text{ and } 2 \quad \text{and} \quad \frac{\hat{\sigma}_{0, n_1, n_2}}{\sigma_{0, n_1, n_2}} \xrightarrow{p} 1.$$

Applying Theorems 1 and 2, under  $H_{0a} : \Sigma_1 = \Sigma_2$ ,

$$(2.7) \quad L_n = \frac{T_{n_1, n_2}}{\hat{\sigma}_{0, n_1, n_2}} \xrightarrow{d} N(0, 1).$$

Hence, the proposed test with a nominal  $\alpha$  level of significance rejects  $H_{0a}$  if  $T_{n_1, n_2} \geq \hat{\sigma}_{0, n_1, n_2} z_\alpha$ , where  $z_\alpha$  is the upper- $\alpha$  quantile of  $N(0, 1)$ .

Let  $\beta_{1, n_1, n_2}(\Sigma_1, \Sigma_2; \alpha) = P(T_{n_1, n_2} / \hat{\sigma}_{0, n_1, n_2} > z_\alpha | H_{1a})$  be the power of the test under  $H_{1a} : \Sigma_1 \neq \Sigma_2$ . From Theorems 1 and 2, the leading order power is

$$(2.8) \quad \Phi \left( -\mathcal{Z}_{n_1, n_2}(\Sigma_1, \Sigma_2) z_\alpha + \frac{\text{tr}\{(\Sigma_1 - \Sigma_2)^2\}}{\sigma_{n_1, n_2}} \right),$$

where  $\mathcal{Z}_{n_1, n_2}(\Sigma_1, \Sigma_2) = (\sigma_{n_1, n_2})^{-1} \{ \frac{2}{n_2} \text{tr}(\Sigma_1^2) + \frac{2}{n_1} \text{tr}(\Sigma_2^2) \}$ . It is the case that  $\mathcal{Z}_{n_1, n_2}(\Sigma_1, \Sigma_2)$  is bounded. To appreciate this, we note that  $\sigma_{n_1, n_2}^2 \geq \frac{4}{n_1^2} \text{tr}^2(\Sigma_1^2) + \frac{4}{n_2^2} \text{tr}^2(\Sigma_2^2)$ . Let  $\gamma_p = \text{tr}(\Sigma_1^2) / \text{tr}(\Sigma_2^2)$  and  $k_n = n_1 / (n_1 + n_2)$ , then

$$\mathcal{Z}_{n_1, n_2}(\Sigma_1, \Sigma_2) \leq \frac{(2/n_2) \text{tr}(\Sigma_1^2) + (2/n_1) \text{tr}(\Sigma_2^2)}{\sqrt{(4/n_1^2) \text{tr}^2(\Sigma_1^2) + (4/n_2^2) \text{tr}^2(\Sigma_2^2)}} =: R_n(\gamma_p),$$

where  $R_n(u) = (\frac{k_n}{1-k_n}u + 1) \{u^2 + (\frac{k_n}{1-k_n})^2\}^{-1/2}$ . Since  $R_n(u)$  is maximized uniquely at  $u^* = (\frac{k_n}{1-k_n})^3$ ,  $\mathcal{Z}_{n_1, n_2}(\Sigma_1, \Sigma_2) \leq \frac{1}{k_n(1-k_n)}$ . Thus,

$$(2.9) \quad \beta_{1, n_1, n_2}(\Sigma_1, \Sigma_2; \alpha) \geq \Phi \left( -\frac{z_\alpha}{k_n(1-k_n)} + \frac{\text{tr}\{(\Sigma_1 - \Sigma_2)^2\}}{\sigma_{n_1, n_2}} \right)$$

implying the power is bounded from below by the probability on the right-hand side.

Both (2.8) and (2.9) indicate that  $\text{SNR}_1(\Sigma_1, \Sigma_2) =: \text{tr}\{(\Sigma_1 - \Sigma_2)^2\} / \sigma_{n_1, n_2}$  is instrumental in determining the power of the test. We term  $\text{SNR}_1(\Sigma_1, \Sigma_2)$  as the signal-to-noise ratio for the current testing problem since  $\text{tr}\{(\Sigma_1 - \Sigma_2)^2\}$  may be viewed as the signal while  $\sigma_{n_1, n_2}$  may be viewed as the level



of the noise. If the signal is strong or the noise is weak so that the signal-to-noise ratio diverges to the infinity, the power will converge to 1. If the signal-to-noise ratio diminishes to 0, the test will not be powerful and cannot distinguish  $H_{0a}$  from  $H_{1a}$ . We note that

$$\sigma_{n_1, n_2}^2 \leq 4 \left\{ \frac{1}{n_1} \text{tr}(\Sigma_1^2) + \frac{1}{n_2} \text{tr}(\Sigma_2^2) \right\}^2 \\ + \max\{8 + 4\Delta_1, 8 + 4\Delta_2\} \left\{ \frac{1}{n_1} \text{tr}(\Sigma_1^2) + \frac{1}{n_2} \text{tr}(\Sigma_2^2) \right\} \text{tr}\{(\Sigma_1 - \Sigma_2)^2\}.$$

Let  $\delta_{1,n} = \{\frac{1}{n_1} \text{tr}(\Sigma_1^2) + \frac{1}{n_2} \text{tr}(\Sigma_2^2)\} / \text{tr}\{(\Sigma_1 - \Sigma_2)^2\}$ , then

$$\text{SNR}_1(\Sigma_1, \Sigma_2) \geq [4\delta_{1,n}^2 + \max\{8 + 4\Delta_1, 8 + 4\Delta_2\}\delta_{1,n}]^{-1/2}.$$

Thus, if the difference between  $\Sigma_1$  and  $\Sigma_2$  is not too small so that

$$(2.10) \quad \begin{aligned} &\text{tr}\{(\Sigma_1 - \Sigma_2)^2\} \text{ is at the same or a larger order} \\ &\text{of } \frac{1}{n_1} \text{tr}(\Sigma_1^2) + \frac{1}{n_2} \text{tr}(\Sigma_2^2), \end{aligned}$$

the test will be powerful. Condition (2.10) is trivially true for fixed-dimensional data while  $n_i \rightarrow \infty$ . For high-dimensional data, it is less automatic as  $\text{tr}(\Sigma_i^2)$  can diverge. To gain further insight on (2.10), let  $\lambda_{i1} \leq \lambda_{i2} \leq \dots \leq \lambda_{ip}$  be the eigenvalues of  $\Sigma_i$ . Then, a sufficient condition for the test to have a nontrivial power is  $\text{tr}\{(\Sigma_1 - \Sigma_2)^2\} = O\{\frac{1}{n_1} \sum_{i=1}^p \lambda_{1i}^2 + \frac{1}{n_2} \sum_{i=1}^p \lambda_{2i}^2\}$ . If all the eigenvalues of  $\Sigma_1$  and  $\Sigma_2$  are bounded away from zero and infinity, (2.10) becomes  $\text{tr}\{(\Sigma_1 - \Sigma_2)^2\} = O(n^{-1}p)$ . Let  $\delta_\beta = p^{-1} \sqrt{\text{tr}\{(\Sigma_1 - \Sigma_2)^2\}}$  be the average signal. Then the test has nontrivial power if  $\delta_\beta$  is at least at the order of  $n^{-1/2}p^{-1/2}$ , which is actually smaller than the conventional order of  $n^{-1/2}$  for fixed-dimension situations. This partially reflects the fact that high data dimensionality is not entirely a curse as there are more data information available as well. If the covariance matrix is believed to have certain structure, for instance banded or bandable in the sense of Bickel and Levina (2008a), we may modify the test statistic so that the comparison of the two covariance matrices is made in the ‘‘important regions’’ under the structure. The modification can be in the form of thresholding, a topic we would not elaborate in this paper; see Cai, Liu and Xia (2011) for research in this direction.

**3. Test for covariance between two sub-vectors.** Let  $X_{ij} = (X_{ij}^{(1)}, X_{ij}^{(2)})$  be a partition of the original data vector into sub-vectors of dimensions of  $p_1$  and  $p_2$ , and  $\Sigma_{i,12} = \text{Cov}(X_{ij}^{(1)}, X_{ij}^{(2)})$  be the covariance between the sub-vectors. The focus in this section is to develop a test procedure for  $H_{0b}: \Sigma_{1,12} = \Sigma_{2,12}$ . Testing for such a hypothesis is importance in its own right, for instance in detecting changes in correlation between two groups of genes under two treatment regimes. It can be also viewed as part of the



effort in reducing the dimensionality in testing high-dimensional variance-covariances. To elaborate on this, consider the partition of  $\Sigma_i$ ,

$$(3.1) \quad \Sigma_i = \begin{pmatrix} \Sigma_{i,11} & \Sigma_{i,12} \\ \Sigma'_{i,12} & \Sigma_{i,22} \end{pmatrix},$$

induced by the partition of the data vectors. Instead of testing on the whole matrices  $\Sigma_1 = \Sigma_2$ , we can first test separately on the two diagonal blocks  $\Sigma_{1,l} = \Sigma_{2,l}$  for  $l = 1$  and  $2$ , by employing the test developed in the previous section based on the sub-vectors of the two sample data respectively. Then, we can test for the off-diagonal blocks  $H_{0b} : \Sigma_{1,12} = \Sigma_{2,12}$  using a test procedure to be developed in this section.

The partition of data vectors also induces a partition of the multivariate model in A3 so that

$$(3.2) \quad X_{ij}^{(1)} = \Gamma_i^{(1)} Z_{ij} + \mu_i^{(1)} \quad \text{and} \quad X_{ij}^{(2)} = \Gamma_i^{(2)} Z_{ij} + \mu_i^{(2)},$$

where  $\Gamma_i^{(1)}$  is  $p_1 \times m_i$  and  $\Gamma_i^{(2)}$  is  $p_2 \times m_i$  such that  $\Gamma'_i = (\Gamma_i^{(1)'} , \Gamma_i^{(2)'})$  and  $\Gamma_i^{(1)} \Gamma_i^{(2)'} = \Sigma_{i,12}$ .

We are interested in testing  $H_{0b} : \Sigma_{1,12} = \Sigma_{2,12}$  vs  $H_{1b} : \Sigma_{1,12} \neq \Sigma_{2,12}$ . The test statistic is aimed at

$$(3.3) \quad \begin{aligned} & \text{tr}\{(\Sigma_{1,12} - \Sigma_{2,12})(\Sigma_{1,12} - \Sigma_{2,12})'\} \\ &= \text{tr}(\Sigma_{1,12} \Sigma'_{1,12}) + \text{tr}(\Sigma_{2,12} \Sigma'_{2,12}) - 2 \text{tr}(\Sigma_{1,12} \Sigma'_{2,12}), \end{aligned}$$

a discrepancy measure between  $\Sigma_{1,12}$  and  $\Sigma_{2,12}$ .

With the same considerations as those when we proposed the estimators in (2.1) and (2.2), we estimate  $\text{tr}(\Sigma_{h,12} \Sigma'_{h,12})$  by

$$(3.4) \quad \begin{aligned} U_{n_h} &= \frac{1}{n_h(n_h - 1)} \sum_{i \neq j} X_{hi}^{(1)'} X_{hj}^{(1)} X_{hj}^{(2)'} X_{hi}^{(2)} \\ &\quad - \frac{2}{n_h(n_h - 1)(n_h - 2)} \sum_{i,j,k}^* X_{hi}^{(1)'} X_{hj}^{(1)} X_{hj}^{(2)'} X_{hk}^{(2)} \\ &\quad + \frac{1}{n_h(n_h - 1)(n_h - 2)(n_h - 3)} \sum_{i,j,k,l}^* X_{hi}^{(1)'} X_{hj}^{(1)} X_{hk}^{(2)'} X_{hl}^{(2)}, \end{aligned}$$

and estimate  $\text{tr}(\Sigma_{1,12} \Sigma'_{2,12})$  by

$$(3.5) \quad \begin{aligned} W_{n_1 n_2} &= \frac{1}{n_1 n_2} \sum_{i,j} X_{1i}^{(1)'} X_{2j}^{(1)} X_{2j}^{(2)'} X_{1i}^{(2)} \\ &\quad - \frac{1}{n_1 n_2 (n_1 - 1)} \sum_{i \neq k, j} X_{1i}^{(1)'} X_{2j}^{(1)} X_{2j}^{(2)'} X_{1k}^{(2)} \end{aligned}$$

$$\begin{aligned}
& - \frac{1}{n_1 n_2 (n_2 - 1)} \sum_{i \neq k, j} X_{2i}^{(1)'} X_{1j}^{(1)} X_{1j}^{(2)'} X_{2k}^{(2)} \\
& + \frac{1}{n_1 n_2 (n_1 - 1)(n_2 - 1)} \sum_{i \neq k, j \neq l} X_{1i}^{(1)'} X_{2j}^{(1)} X_{1k}^{(2)'} X_{2l}^{(2)}.
\end{aligned}$$

Both  $U_{nh}$  and  $W_{n_1 n_2}$  are linear combinations of U-statistics.

Combining these estimators together leads to an unbiased estimator of  $\text{tr}\{(\Sigma_{1,12} - \Sigma_{2,12})(\Sigma_{1,12} - \Sigma_{2,12})'\}$ ,

$$(3.6) \quad S_{n_1, n_2} = U_{n_1} + U_{n_2} - 2W_{n_1 n_2},$$

which is also invariant under the location shift and orthogonal rotations.

To establish the asymptotic normality of  $S_{n_1, n_2}$ , we need an extra assumption regarding the off-diagonal sub-matrices.

A4. As  $\min\{n_1, n_2\} \rightarrow \infty$ , for any  $i, j, k$  and  $l \in \{1, 2\}$ .

$$(3.7) \quad \text{tr}(\Sigma_{i,11} \Sigma_{j,12} \Sigma_{k,22} \Sigma_{l,12}') = o\{\text{tr}(\Sigma_{i,11} \Sigma_{j,11}) \text{tr}(\Sigma_{k,22} \Sigma_{l,22})\}.$$

Derivations leading to (6.5) in Section 6 show that, under A2, A3 and A4, the leading order variance of  $S_{n_1, n_2}$  is

$$\begin{aligned}
\omega_{n_1, n_2}^2 = & \sum_{i=1}^2 \left[ \frac{2}{n_i^2} \text{tr}^2(\Sigma_{i,12} \Sigma_{i,12}') + \frac{2}{n_i^2} \text{tr}(\Sigma_{i,11}^2) \text{tr}(\Sigma_{i,22}^2) \right. \\
& + \frac{4}{n_i} \text{tr}\{(\Sigma_{i,12} \Sigma_{1,12}' - \Sigma_{i,12} \Sigma_{2,12}')^2\} \\
(3.8) \quad & + \frac{4}{n_i} \text{tr}\{(\Sigma_{i,11} \Sigma_{1,12} - \Sigma_{i,11} \Sigma_{2,12})(\Sigma_{i,22} \Sigma_{1,12}' - \Sigma_{i,22} \Sigma_{2,12}')\} \\
& + \left. \frac{4\Delta_i}{n_i} \text{tr}\{\Gamma_i^{(1)'}(\Sigma_{1,12} - \Sigma_{2,12})\Gamma_i^{(2)} \circ \Gamma_i^{(1)'}(\Sigma_{1,12} - \Sigma_{2,12})\Gamma_i^{(2)}\} \right] \\
& + \frac{4}{n_1 n_2} \text{tr}^2(\Sigma_{1,12} \Sigma_{2,12}') + \frac{4}{n_1 n_2} \text{tr}(\Sigma_{1,11} \Sigma_{2,11}) \text{tr}(\Sigma_{1,22} \Sigma_{2,22}).
\end{aligned}$$

Similarly to the analysis on  $T_{n_1, n_2}$  in the previous section, the asymptotic normality of  $S_{n_1, n_2}$  can be established in the following theorem.

**THEOREM 3.** *Under Conditions A1–A4, as  $\min\{n_1, n_2\} \rightarrow \infty$ ,*

$$\omega_{n_1, n_2}^{-1} [S_{n_1, n_2} - \text{tr}\{(\Sigma_{1,12} - \Sigma_{2,12})(\Sigma_{1,12} - \Sigma_{2,12})'\}] \xrightarrow{d} N(0, 1).$$

Under  $H_{0b} : \Sigma_{1,12} = \Sigma_{2,12} = \Sigma_{12}$ , say,  $\omega_{n_1, n_2}^2$  becomes

$$\begin{aligned}
\omega_{0, n_1, n_2}^2 = & 2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)^2 \text{tr}^2(\Sigma_{12} \Sigma_{12}') + 2 \sum_{i=1}^2 \frac{1}{n_i^2} \text{tr}(\Sigma_{i,11}^2) \text{tr}(\Sigma_{i,22}^2) \\
(3.9) \quad & + \frac{4}{n_1 n_2} \text{tr}(\Sigma_{1,11} \Sigma_{2,11}) \text{tr}(\Sigma_{1,22} \Sigma_{2,22}).
\end{aligned}$$

In order to formulate a test procedure,  $\omega_{0,n_1,n_2}^2$  needs to be estimated. An unbiased estimator of  $\text{tr}(\Sigma_{h,ll}^2)$  for  $h = 1$  or  $2$  and  $l = 1$  or  $2$ , is

$$\begin{aligned} A_{n_h}^{(l)} &= \frac{1}{n_h(n_h - 1)} \sum_{i \neq j} (X_{hi}^{(l)'} X_{hj}^{(l)})^2 - \frac{2}{n_h(n_h - 1)(n_h - 2)} \sum_{i,j,k}^* X_{hi}^{(l)'} X_{hj}^{(l)} X_{hj}^{(l)'} X_{hk}^{(l)} \\ &\quad + \frac{1}{n_h(n_h - 1)(n_h - 2)(n_h - 3)} \sum_{i,j,k,l}^* X_{hi}^{(l)'} X_{hj}^{(l)} X_{hk}^{(l)'} X_{hl}^{(l)}. \end{aligned}$$

Similarly, an unbiased estimator of  $\text{tr}(\Sigma_{1,hh}\Sigma_{2,hh})$ , for  $h = 1$  or  $2$ , is

$$\begin{aligned} C_{n_1 n_2}^{(h)} &= \frac{1}{n_1 n_2} \sum_{i,j} (X_{1i}^{(h)'} X_{2j}^{(h)})^2 - \frac{1}{n_1 n_2 (n_1 - 1)} \sum_{i \neq k, j} X_{1i}^{(h)'} X_{2j}^{(h)} X_{2j}^{(h)'} X_{1k}^{(h)} \\ &\quad - \frac{1}{n_1 n_2 (n_2 - 1)} \sum_{i \neq k, j} X_{2i}^{(h)'} X_{1j}^{(h)} X_{1j}^{(h)'} X_{2k}^{(h)} \\ &\quad + \frac{1}{n_1 n_2 (n_1 - 1)(n_2 - 1)} \sum_{i \neq k, j \neq l} X_{1i}^{(h)'} X_{2j}^{(h)} X_{1k}^{(h)'} X_{2l}^{(h)}. \end{aligned}$$

Then under  $H_{0b}$ , an unbiased estimator of  $\omega_{0,n_1,n_2}^2$  is

$$\hat{\omega}_{0,n_1,n_2}^2 = 2 \left( \frac{U_{n_1}}{n_2} + \frac{U_{n_2}}{n_1} \right)^2 + \frac{2}{n_1^2} A_{n_1}^{(1)} A_{n_1}^{(2)} + \frac{2}{n_2^2} A_{n_2}^{(1)} A_{n_2}^{(2)} + \frac{4}{n_1 n_2} C_{n_1 n_2}^{(1)} C_{n_1 n_2}^{(2)}.$$

The following theorem shows that  $\hat{\omega}_{0,n_1,n_2}^2$  is ratio-consistent to  $\omega_{0,n_1,n_2}^2$ .

**THEOREM 4.** *Under Conditions A1–A4, and  $H_{0b} : \Sigma_{1,12} = \Sigma_{2,12}$ ,*

$$\frac{\hat{\omega}_{0,n_1,n_2}^2}{\omega_{0,n_1,n_2}^2} \xrightarrow{p} 1.$$

Applying Theorems 3 and 4, we have, under  $H_{0b}$ ,

$$\frac{S_{n_1,n_2}}{\hat{\omega}_{0,n_1,n_2}} \xrightarrow{d} N(0, 1).$$

This suggests an  $\alpha$ -level test that rejects  $H_{0b}$  if  $S_{n_1,n_2} \geq \hat{\omega}_{0,n_1,n_2} z_\alpha$ . The power of the proposed test under  $H_{1b} : \Sigma_{1,12} \neq \Sigma_{2,12}$  is

$$\beta_{2,n_1,n_2}(\Sigma_{1,12}, \Sigma_{2,12}; \alpha) = P(S_{n_1,n_2} / \hat{\omega}_{0,n_1,n_2} > z_\alpha | H_{1b}).$$

From Theorems 3 and 4, the leading order power is

$$\Phi \left( -\frac{\tilde{\omega}}{\omega_{n_1,n_2}} z_\alpha + \frac{\text{tr}\{(\Sigma_{1,12} - \Sigma_{2,12})(\Sigma_{1,12} - \Sigma_{2,12})'\}}{\omega_{n_1,n_2}} \right),$$

where

$$\begin{aligned}\tilde{\omega}^2 = & 2 \left\{ \frac{\text{tr}(\Sigma_{1,12} \Sigma'_{1,12})}{n_2} + \frac{\text{tr}(\Sigma_{2,12} \Sigma'_{2,12})}{n_1} \right\}^2 + \frac{2}{n_1^2} \text{tr}(\Sigma_{1,11}^2) \text{tr}(\Sigma_{1,22}^2) \\ & + \frac{2}{n_2^2} \text{tr}(\Sigma_{2,11}^2) \text{tr}(\Sigma_{2,22}^2) + \frac{4}{n_1 n_2} \text{tr}(\Sigma_{1,11} \Sigma_{2,11}) \text{tr}(\Sigma_{1,22} \Sigma_{2,22}).\end{aligned}$$

Let  $\eta_p = \text{tr}(\Sigma_{1,12} \Sigma'_{1,12}) / \text{tr}(\Sigma_{2,12} \Sigma'_{2,12})$ . It may be shown that

$$\frac{\tilde{\omega}}{\omega_{n_1, n_2}} \leq \sqrt{R^2(\eta_p) + 1},$$

where  $R(\gamma_p)$  is the same function defined in Section 2. Hence, asymptotically,

$$\begin{aligned}\beta_{2, n_1, n_2}(\Sigma_{1,12}, \Sigma_{2,12}; \alpha) \\ \geq \Phi \left( -\frac{z_\alpha \sqrt{1 + k_n^2(1 - k_n)^2}}{k_n(1 - k_n)} + \frac{\text{tr}\{(\Sigma_{1,12} - \Sigma_{2,12})(\Sigma_{1,12} - \Sigma_{2,12})'\}}{\omega_{n_1, n_2}} \right).\end{aligned}$$

This implies that

$$\text{SNR}_2 =: \text{tr}\{(\Sigma_{1,12} - \Sigma_{2,12})(\Sigma_{1,12} - \Sigma_{2,12})'\} / \omega_{n_1, n_2}$$

is the key quantity that determines the power of the test. Furthermore, let

$$\delta_{2, n} = \frac{(1/n_1) \text{tr}(\Sigma_{1,11}) \text{tr}(\Sigma_{1,22}) + (1/n_2) \text{tr}(\Sigma_{2,11}) \text{tr}(\Sigma_{2,22})}{\text{tr}\{(\Sigma_{1,12} - \Sigma_{2,12})(\Sigma_{1,12} - \Sigma_{2,12})'\}}.$$

It can be shown that

$$(3.10) \quad \text{SNR}_2 \geq [4\delta_{2, n}^2 + \max\{8 + 4\Delta_1, 8 + 4\Delta_2\} \delta_{2, n}]^{-1/2}.$$

Hence, the test is powerful if the difference between  $\Sigma_{1,12}$  and  $\Sigma_{2,12}$  is not too small so that  $\text{tr}\{(\Sigma_{1,12} - \Sigma_{2,12})(\Sigma_{1,12} - \Sigma_{2,12})'\}$  is at the order of  $\sum_{i=1}^2 \frac{1}{n_i} \times \text{tr}(\Sigma_{i,11}) \text{tr}(\Sigma_{i,22})$  or larger. A further analysis on the power, similar to that given at the end of last section, can be made. Here for the sake of brevity, we will not report.

**4. Simulation studies.** We report results from simulation experiments which were designed to evaluate the performance of the two proposed tests. A range of dimensionality and sample sizes was considered which allowed  $p$  to increase as the sample sizes were increased. This was designed to confirm the asymptotic results reported in the previous sections.

We first considered the test for  $H_{0a} : \Sigma_1 = \Sigma_2$  regarding the whole variance-covariance matrices. To compare with the conventional likelihood ratio (LR) test and the corrected LR test proposed by Bai et al. (2009), we first considered cases of  $p \leq \min\{n_1, n_2\}$  and the normally distributed data. Specifically, to create the null hypothesis, we simulated both samples from the

TABLE 1

*Empirical sizes and powers of the conventional likelihood ratio (LR), the corrected likelihood ratio (CLR) and the proposed tests (Proposed) for the variance-covariance, based on 1000 replications with normally distributed  $\{Z_{ijk}\}$*

$(p, n_1, n_2)$	Methods	Size	Power		
			$\theta_1 = 0.5$	$\theta_1 = 0.3$	$\theta_1 = 0.2$
(40, 60, 60)	LRT	1	1	1	1
	CLRT	0.043	0.999	0.509	0.172
	Proposed	0.052	0.999	0.734	0.271
(80, 120, 120)	LRT	1	1	1	1
	CLRT	0.045	1	0.946	0.421
	Proposed	0.053	1	0.997	0.713
(120, 180, 180)	LRT	1	1	1	1
	CLRT	0.062	1	1	0.713
	Proposed	0.045	1	1	0.958

$p$ -dimensional standard normal distribution. To evaluate the power of the three tests, we set the first population to be the  $p$ -dimensional standard normally distributed while simulating the second population according to

$$(4.1) \quad X_{ijk} = Z_{ijk} + \theta_1 Z_{ijk+1},$$

where  $\{Z_{ijk}\}$  were i.i.d. standard normally distributed, and  $\theta_1 = 0.5, 0.3$  and  $0.2$ , respectively. As  $\theta_1$  was decreased, the signal strength for the test became weaker. We chose  $(p, n_1, n_2) = (40, 60, 60), (80, 120, 120)$  and  $(120, 180, 180)$ , respectively. The empirical size and power for the three tests are reported in Table 1. All the simulation results reported in this section were based on 1000 simulations with the nominal significance level to be 5%.

We then carried out simulations for situations where  $p$  was much larger than the sample sizes. In this case, only the proposed test was considered, as both the LR and the corrected LR tests were no longer applicable. We chose a set of data dimensions from 32 to 700, while the sample sizes ranged from 20 to 100, respectively. We considered the moving average model (4.1) with  $\theta_1 = 2$  as the null model of both populations for size evaluation. To assess the power performance, the first population was generated according to (4.1), while the second was from

$$(4.2) \quad X_{ijk} = Z_{ijk} + \theta_1 Z_{ijk+1} + \theta_2 Z_{ijk+2},$$

where  $\theta_1 = 2$  and  $\theta_2 = 1$ . Three combinations of distributions were experimented for the i.i.d. sequences  $\{Z_{ijk}\}_{k=1}^p$  in models (4.1) and (4.2), respectively. They were: (i) both sequences were the standard normal; (ii) the centralized Gamma(4, 0.5) for Sample 1 and the centralized Gamma(0.5,  $\sqrt{2}$ ) for Sample 2; (iii) the standard normal for Sample 1 and the centralized

TABLE 2

*Empirical sizes and powers of the proposed test for the variance-covariance matrices, based on 1000 replications with normally distributed  $\{Z_{ijk}\}$  in Models (4.1) and (4.2)*

$n_1 = n_2$	$p$					
	32	64	128	256	512	700
Sizes						
20	0.044	0.054	0.051	0.048	0.051	0.038
50	0.052	0.060	0.033	0.043	0.054	0.049
80	0.054	0.060	0.047	0.048	0.052	0.053
100	0.056	0.049	0.052	0.046	0.049	0.048
Powers						
20	0.291	0.256	0.267	0.277	0.282	0.291
50	0.746	0.821	0.830	0.837	0.832	0.849
80	0.957	0.992	0.991	0.998	0.999	0.998
100	0.994	1	0.999	1	1	1

Gamma(0.5,  $\sqrt{2}$ ) for Sample 2. The last two combinations were designed to assess the performance under nonnormality. The empirical size and power of the test are reported in Tables 2–4.

We observed from Table 1 that the size of the conventional LR test was grossly distorted, confirming its breakdown under even mild dimensionality, discovered in Bai et al. (2009). The severely distorted size for the LR test made its power artificially high. Both the corrected LR test and the proposed test had quite accurate size approximation to the nominal 5% level for all

TABLE 3

*Empirical sizes and powers of the proposed test for the variance-covariance matrices, based on 1000 replications with Gamma distributed  $\{Z_{ijk}\}$  in Models (4.1) and (4.2)*

$n_1 = n_2$	$p$					
	32	64	128	256	512	700
Sizes						
20	0.119	0.117	0.069	0.063	0.051	0.040
50	0.150	0.110	0.094	0.052	0.053	0.051
80	0.155	0.111	0.093	0.067	0.064	0.044
100	0.148	0.120	0.084	0.056	0.058	0.053
Powers						
20	0.299	0.282	0.290	0.309	0.265	0.277
50	0.574	0.665	0.693	0.750	0.801	0.828
80	0.804	0.886	0.942	0.968	0.991	0.986
100	0.899	0.945	0.986	0.995	0.998	1

TABLE 4

*Empirical sizes and powers of the proposed test for the variance-covariance matrices, based on 1000 replications with the mixed normal and Gamma distributions for  $\{Z_{ijk}\}$  in Models (4.1) and (4.2)*

$n_1 = n_2$	$p$					
	32	64	128	256	512	700
Sizes						
20	0.108	0.099	0.076	0.059	0.070	0.050
50	0.117	0.111	0.069	0.068	0.057	0.053
80	0.124	0.099	0.091	0.065	0.064	0.060
100	0.150	0.122	0.085	0.069	0.056	0.047
Powers						
20	0.256	0.296	0.278	0.297	0.276	0.295
50	0.606	0.659	0.724	0.766	0.824	0.823
80	0.805	0.890	0.950	0.977	0.989	0.992
100	0.904	0.958	0.982	0.996	0.999	1

cases in Table 1. Both tests enjoyed perfect power at  $\theta_1 = 0.5$ , when the signal strength of the tests was strong. When the value of  $\theta_2$  decreased, both tests had smaller power, although the proposed test was slightly more powerful than the corrected LR test at  $\theta_1 = 0.3$  and much more so at  $\theta_1 = 0.2$ , when the signal strength was weaker.

The simulation results for the proposed test with dimensions much larger than the sample sizes and for nonnormally distributed data are reported in Tables 2–4. We note that the LR tests are not applicable for the setting. The simulation results show that the proposed test had quite accurate and robust size approximation in a quite wider range of dimensionality and distributions, considered in the simulation experiments. The tables also show that the power of the proposed tests was quite satisfactory and was increased as the dimension and the sample sizes became larger.

We then conducted simulations to evaluate the performance of the second test for  $H_{0b} : \Sigma_{1,12} = \Sigma_{2,12}$ . We partition equally the entire random vector  $X_{ij}$  into two subvectors of  $p_1 = p/2$  and  $p_2 = p - p_1$ . To ensure sufficient number of nonzero elements in the off-diagonal sub-matrices  $\Sigma_{1,12}$  and  $\Sigma_{2,12}$  when the dimension was increased, we considered a moving average model of order  $m_1$ , which is much larger than the orders used in (4.1) and (4.2). In the size evaluation,

$$(4.3) \quad X_{ijk} = Z_{ijk} + \alpha_1 Z_{ijk+1} + \cdots + \alpha_{m_1} Z_{ijk+m_1}$$

for  $i = 1, 2, j = 1, \dots, n_i$ , where all the  $\alpha_i$  coefficients were chosen to be 0.1. In the simulation for the power, we generated the first sample according to



TABLE 5  
*Empirical sizes and powers of the proposed test for the covariance between two sub-vectors, based on 1000 replications for normally distributed  $\{Z_{ijk}\}$  in Models (4.3) and (4.4)*

$n_1 = n_2$	$p$				
	50	100	200	500	700
Sizes					
20	0.069	0.071	0.070	0.065	0.077
50	0.064	0.056	0.064	0.063	0.055
80	0.057	0.046	0.056	0.073	0.052
100	0.047	0.062	0.055	0.054	0.048
Powers					
20	0.639	0.625	0.628	0.620	0.615
50	0.993	0.994	0.982	0.983	0.989
80	1	1	1	1	1
100	1	1	1	1	1

the above (4.3) and the second from

$$(4.4) \quad X_{ijk} = Z_{ijk} + \beta_1 Z_{ijk+1} + \cdots + \beta_{m_2} Z_{ijk+m_2}$$

for  $j = 1, \dots, n_2$ , where the  $\beta_i$  were chosen to be 0.8. We chose the lengths of the moving average  $m_1$  and  $m_2$  according to the dimension  $p$  such that as  $p$  was increased, the values of  $m_1$  and  $m_2$  were increased as well. Specifically, we set  $(m_1, m_2, p) = (2, 25, 50), (3, 50, 100), (7, 100, 200), (12, 250, 500)$  and  $(18, 300, 700)$ , respectively. Two distributions were considered for the i.i.d. sequences  $\{Z_{ijk}\}_{k=1}^p$  in (4.3) and (4.4): (i) both sequences were standard normally distributed; (ii) the centralized Gamma(4, 0.5) for Sample 1 and the centralized Gamma(0.5,  $\sqrt{2}$ ) for Sample 2. The simulation results for the second test are reported in Table 5 for the normally distributed case and Table 6 for the Gamma distributed case.

We observed from Table 5 that the empirical sizes of the proposed test converged to the nominal 5% quite rapidly, while the powers were quite high and quickly increased to 1. For the Gamma distributed case reported in Table 6, the convergence of the empirical sizes to the nominal level was slower than the normally distributed case indicating that the convergence of the asymptotic normality depends on the underlying distribution, as well as the sample size and dimensionality. The powers in Table 6 were reasonable, although they were smaller than the corresponding normally distributed case in Table 5. Nevertheless, the power was quite responsive to the increase of  $p$  and the sample sizes.

**5. An empirical study.** We report an empirical study on a leukemia data by applying the proposed tests on the variance–covariance matrices. The da-

TABLE 6  
*Empirical sizes and powers of the proposed test for the covariances between two sub-vectors, based on 1000 replications with Gamma distributed  $\{Z_{ijk}\}$  in Models (4.3) and (4.4)*

$n_1 = n_2$	$p$				
	50	100	200	500	700
Sizes					
20	0.105	0.092	0.085	0.082	0.082
50	0.101	0.090	0.081	0.088	0.090
80	0.107	0.094	0.083	0.078	0.065
100	0.093	0.083	0.093	0.059	0.071
Powers					
20	0.499	0.501	0.519	0.482	0.502
50	0.775	0.802	0.783	0.754	0.777
80	0.945	0.923	0.921	0.922	0.923
100	0.974	0.957	0.969	0.964	0.960

ta [Chiaretti et al. (2004)], available from <http://www.bioconductor.org/>, consist of microarray expressions of 128 patients with either T-cell or B-cell acute lymphoblastic leukemia (ALL); see Dudoit, Keles and van der Laan (2008) and Chen and Qin (2010) for analysis on the same dataset. We considered a subset of the ALL data of 79 patients with the B-cell ALL. We were interested in two types of the B-cell tumors: BCR/ABL, one of the most frequent cytogenetic abnormalities in human leukemia, and NEG, the cytogenetically normal B-cell ALL. The number of patients with BCR/ABL was 37 and that with NEG was 42.

A major motivation for developing the proposed test procedures for high-dimensional variance–covariance matrices comes from the need to identify sets of genes which are significantly different with respect to two treatments in genetic research; see Barry, Nobel and Wright (2005), Efron and Tibshirini (2007), Newton et al. (2007) and Nettleton, Recknor and Reecy (2008) for comprehensive discussions. Biologically speaking, each gene does not function individually, but rather tends to work with others to achieve certain biological tasks. Gene-sets are technically defined vocabularies which produce names of gene-sets (also called GO terms). There are three categories of Gene ontologies of interest: Biological Processes (BP), Cellular Components (CC) and Molecular Functions (MF). For the ALL data, a preliminary screening with gene-filtering left a total number of 2391 genes for analysis with 1599 unique GO terms in BP category, 290 in CC and 357 in MF.

Let us denote  $\mathcal{S}_1, \dots, \mathcal{S}_q$  for  $q$  gene-sets, where  $\mathcal{S}_g$  consists of  $p_g$  genes. Let  $F_{1\mathcal{S}_g}$  and  $F_{2\mathcal{S}_g}$  be the distribution functions corresponding to  $\mathcal{S}_g$  under the treatment and control, and  $\mu_{1\mathcal{S}_g}$  and  $\mu_{2\mathcal{S}_g}$  be their respective means,

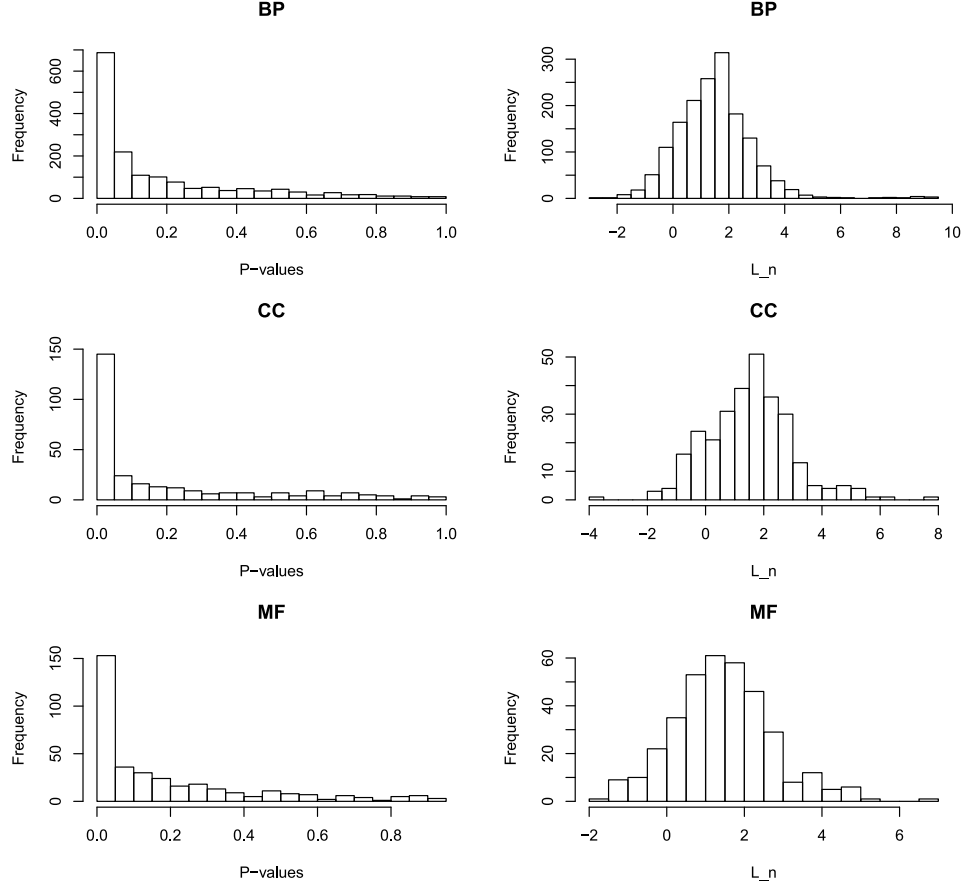


FIG. 1. Histograms of  $p$ -values (left panels) for testing two covariance matrices and test statistic  $L_n$  (right panels) for the three gene-categories.

and  $\Sigma_{1\mathcal{S}_g}$  and  $\Sigma_{2\mathcal{S}_g}$  be their respective variance-covariance matrices. Our first hypotheses of interest are  $H_{0g}: \Sigma_{1\mathcal{S}_g} = \Sigma_{2\mathcal{S}_g}$  for  $g = 1, \dots, q$  regarding the variance-covariance matrices. For the second hypothesis, we divided each gene-set into two sub-vectors by selecting the first  $\lfloor p/2 \rfloor$  dimensions of the gene-set as the first segment and the rest as the second.

We first applied the proposed test for the equality of the entire variance-covariance matrices and obtained the  $p$ -value for each gene-set. The  $p$ -values and the values of the test statistics  $L_n$  as given in (2.7) are displayed in Figure 1 for the three gene-categories. By controlling the false discovery rate [FDR, Benjamini and Hochberg (1995)] at 0.05, 338 GO terms were declared significant in the BP category, 77 in the CC and 75 in the MF, indicating that the dependence structure among the gene-sets was significantly different between the BCR/ABL and the NEG ALL patients for a large number

TABLE 7

*Number of GO terms which were tested significantly different at the diagonal blocks, off-diagonal blocks and both diagonal and off-diagonal blocks, respectively*

	Diagonal only	Off-diagonal only	Both	Total
BP	115	17	206	338
CC	26	1	50	77
MF	22	0	53	75

of gene sets. That a relatively large number of gene-sets being declared significant by the proposed test was not entirely surprising, as we observe from Figure 1 that there were very large number of  $p$ -values which were very close to 0.

For those GO terms which had been declared having different variance-covariance matrices, we carried out a follow-up analysis trying to gain more details on the differences by partitioning the variance-covariance into four blocks in the form of (3.1) with  $p_1 = \lfloor p/2 \rfloor$  and  $p_2 = p - p_1$ . We want to know if the difference was caused by the diagonal blocks or the off-diagonal blocks. The tests on the two diagonal blocks were conducted using the first proposed test for the variance-covariance matrix but with  $p_1$  or  $p_2$  dimensions, respectively. The tests on the off-diagonal blocks were conducted by employing the second proposed test for covariances between the two sub-vectors. The results are summarized in Table 7, which provides the numbers of gene-sets which were tested significant in the diagonal matrices only, the off-diagonal matrix only, and both at 5%. There were far more gene-sets which had both diagonal and off-diagonal matrices being significantly different, and it was less likely that the off-diagonal matrices were different while the diagonal matrices were otherwise. It was a little surprising to see that the numbers of significant gene-sets for the diagonally-only, off-diagonal only and both in each functional category added up to the total numbers exactly for all three gene-categories.

As we have stated in the Introduction, the proposed tests are part of the effort in testing for high-dimensional distributions between two treatments. However, directly testing on the distribution functions is quite challenging due to the high dimensionality as such tests may endure low power. A realistic and intuitive way is to test for simpler characteristics of the distributions, for instance testing for the means as in Bai and Saranadasa (1996) and Chen and Qin (2010), and the variance-covariance as considered in this paper. For the ALL data, in addition to testing for the variance-covariance, we also carried out tests for the means proposed in Chen and Qin (2010) at a level of 5%. Table 8 contains two by two classifications on the number and the probability of gene-sets which are rejected/not rejected by the tests for the mean

TABLE 8

*Two by two classifications on the number (probability) of GO-terms rejected/not rejected by the tests for the means and the variances for the three functional categories, respectively*

Variance test	Mean test	
	Rejected	Not rejected
(a) Biological Processes (BP)		
Rejected	314 (0.196)	22 (0.015)
Not rejected	1000 (0.625)	263 (0.164)
(b) Cellular Components (CC)		
Rejected	77 (0.266)	4 (0.014)
Not rejected	164 (0.566)	45 (0.154)
(c) Molecular Functions (MF)		
Rejected	86 (0.241)	1 (0.003)
Not rejected	203 (0.568)	67 (0.188)

and the variance respectively. It is observed that it is far more likely for the means to be significantly different than the variance-covariance, with the probability of rejection being around 0.8 for the means versus 0.2 to 0.3 for the covariance for the three functional categories. Given a gene-set which was not tested significant for the means, the conditional probability of being tested significant for the covariance is lower than that given a gene-set was not tested significant for the means. These were confirmed by conducting the chi-square test for association for the three gene-set categories, which rejected overwhelmingly (with  $p$ -values all less than 0.0005) the hypothesis of no-association between being tested significant for the mean and the variance. For this particular dataset, the tests for the means were quite effective in disclosing most of the differentially expressed gene-sets. However, we do see that for Biological Processes and Cellular Component categories, among those whose means were not declared significantly different, there were about 10% of gene-sets having significant different covariance structures.

**6. Technical details.** As both  $T_{n_1, n_2}$  and  $S_{n_1, n_2}$  are invariant under the location transformation, we assume  $\mu_i = 0$  throughout this section.

6.1. *Derivations of  $\text{Var}(T_{n_1, n_2})$  and  $\text{Var}(S_{n_1, n_2})$ .* Recall that  $T_{n_1, n_2} = A_{n_1} + A_{n_2} - 2C_{n_1 n_2}$ . It is straightforward to show that  $E(T_{n_1, n_2}) = \text{tr}\{(\Sigma_1 - \Sigma_2)^2\}$ . By noticing that  $\text{Cov}(A_{n_1}, A_{n_2}) = 0$ ,

$$\begin{aligned}
 \text{Var}(T_{n_1, n_2}) &= \text{Var}(A_{n_1}) + \text{Var}(A_{n_2}) + 4 \text{Var}(C_{n_1 n_2}) \\
 (6.1) \quad &\quad - 4 \text{Cov}(A_{n_1}, C_{n_1 n_2}) - 4 \text{Cov}(A_{n_2}, C_{n_1 n_2}).
 \end{aligned}$$

Adopting results from Chen, Zhang and Zhong (2010), for  $h = 1$  or  $2$ ,

$$(6.2) \quad \begin{aligned} \text{Var}(A_{n_h}) &= \frac{4}{n_h^2} \text{tr}^2(\Sigma_h^2) + \frac{8}{n_h} \text{tr}(\Sigma_h^4) + \frac{4\Delta_h}{n_h} \text{tr}(\Gamma'_h \Gamma_h \Gamma'_h \Gamma_h \circ \Gamma'_h \Gamma_h \Gamma'_h \Gamma_h) \\ &\quad + O\left\{ \frac{1}{n_h^3} \text{tr}^2(\Sigma_h^2) + \frac{1}{n_h^2} \text{tr}(\Sigma_h^4) \right\}. \end{aligned}$$

Furthermore, we obtain

$$(6.3) \quad \begin{aligned} \text{Var}(C_{n_1 n_2}) &= \frac{2}{n_1 n_2} \text{tr}^2(\Sigma_1 \Sigma_2) + \left( \frac{2}{n_1} + \frac{2}{n_2} \right) \text{tr}(\Sigma_1 \Sigma_2 \Sigma_1 \Sigma_2) \\ &\quad + \frac{\Delta_1}{n_1} \text{tr}(\Gamma'_1 \Gamma_2 \Gamma'_2 \Gamma_1 \circ \Gamma'_1 \Gamma_2 \Gamma'_2 \Gamma_1) \\ &\quad + \frac{\Delta_2}{n_2} \text{tr}(\Gamma'_2 \Gamma_1 \Gamma'_1 \Gamma_2 \circ \Gamma'_2 \Gamma_1 \Gamma'_1 \Gamma_2) + o\left\{ \frac{1}{n_1 n_2} \text{tr}^2(\Sigma_1 \Sigma_2) \right\} \\ &\quad + O\left[ \left\{ \frac{1}{\sqrt{n_1 n_2}} + \frac{1}{n_1 n_2} + \sum_{i=1}^2 \left( \frac{1}{\sqrt{n_i}} + \frac{1}{n_i} \right) \right\} \text{Var}(C_{n_1 n_2, 1}) \right]. \end{aligned}$$

By carrying out similar procedures, we are able to obtain  $\text{Cov}(A_{n_1}, C_{n_1 n_2})$  and  $\text{Cov}(A_{n_2}, C_{n_1 n_2})$ . After we substitute all the results into (6.1),

$$(6.4) \quad \begin{aligned} \text{Var}(T_{n_1 n_2}) &= \sum_{i=1}^2 \left[ \frac{4}{n_i^2} \text{tr}^2(\Sigma_i^2) + \frac{8}{n_i} \text{tr}(\Sigma_i^4) + \frac{4\Delta_i}{n_i} \text{tr}(\Gamma'_i \Gamma_i \Gamma'_i \Gamma_i \circ \Gamma'_i \Gamma_i \Gamma'_i \Gamma_i) \right. \\ &\quad \left. - \frac{16}{n_i} \text{tr}(\Sigma_i^2 \Sigma_1 \Sigma_2) - \frac{8\Delta_i}{n_i} \text{tr}(\Gamma'_i \Sigma_1 \Gamma_i \circ \Gamma'_i \Sigma_2 \Gamma_i) \right] \\ &\quad + \frac{8}{n_1 n_2} \text{tr}^2(\Sigma_1 \Sigma_2) + \left( \frac{8}{n_1} + \frac{8}{n_2} \right) \text{tr}(\Sigma_1 \Sigma_2 \Sigma_1 \Sigma_2) \\ &\quad + \frac{4\Delta_1}{n_1} \text{tr}(\Gamma'_1 \Gamma_2 \Gamma'_2 \Gamma_1 \circ \Gamma'_1 \Gamma_2 \Gamma'_2 \Gamma_1) \\ &\quad + \frac{4\Delta_2}{n_2} \text{tr}(\Gamma'_2 \Gamma_1 \Gamma'_1 \Gamma_2 \circ \Gamma'_2 \Gamma_1 \Gamma'_1 \Gamma_2) \\ &\quad + o\left\{ \frac{1}{n_1 n_2} \text{tr}^2(\Sigma_1 \Sigma_2) \right\} \\ &\quad + O\left[ \left\{ \frac{1}{\sqrt{n_1 n_2}} + \frac{1}{n_1 n_2} + \sum_{i=1}^2 \left( \frac{1}{\sqrt{n_i}} + \frac{1}{n_i} \right) \right\} \text{Var}(C_{n_1 n_2, 1}) \right. \\ &\quad \left. + \sum_{i=1}^2 \left\{ \frac{1}{n_i^2} \text{tr}(\Sigma_i^4) + \frac{1}{n_i^3} \text{tr}^2(\Sigma_i^2) \right\} \right]. \end{aligned}$$

Similarly to  $T_{n_1, n_2}$ , we have  $E(S_{n_1, n_2}) = \text{tr}\{(\Sigma_{1,12} - \Sigma_{2,12})(\Sigma_{1,12} - \Sigma_{2,12})'\}$  and the leading order terms in  $\text{Var}(S_{n_1, n_2})$  are given by

$$\begin{aligned}
 \text{Var}(S_{n_1, n_2}) = & \sum_{i=1}^2 \left[ \frac{2}{n_i^2} \text{tr}^2(\Sigma_{i,12} \Sigma'_{i,12}) + \frac{2}{n_i^2} \text{tr}(\Sigma_{i,11}^2) \text{tr}(\Sigma_{i,22}^2) \right. \\
 & + \frac{4}{n_i} \text{tr}\{(\Sigma_{i,12} \Sigma'_{1,12} - \Sigma_{i,12} \Sigma'_{2,12})^2\} \\
 & + \frac{4}{n_i} \text{tr}\{(\Sigma_{i,11} \Sigma_{1,12} - \Sigma_{i,11} \Sigma_{2,12})(\Sigma_{i,22} \Sigma'_{1,12} - \Sigma_{i,22} \Sigma'_{2,12})\} \\
 & + \frac{4\Delta_i}{n_i} \\
 & \left. \times \text{tr}\{\Gamma_i^{(1)'}(\Sigma_{1,12} - \Sigma_{2,12})\Gamma_i^{(2)} \circ \Gamma_i^{(1)'}(\Sigma_{1,12} - \Sigma_{2,12})\Gamma_i^{(2)}\} \right] \\
 & + \frac{4}{n_1 n_2} \text{tr}^2(\Sigma_{1,12} \Sigma'_{2,12}) + \frac{4}{n_1 n_2} \text{tr}(\Sigma_{1,11} \Sigma_{2,11}) \text{tr}(\Sigma_{1,22} \Sigma_{2,22}).
 \end{aligned} \tag{6.5}$$

6.2. *Proof of Theorem 1.* The leading order terms in  $\text{Var}(T_{n_1, n_2})$  are contributed by  $A_{n_h, 1}$  for  $h = 1, 2$  and  $C_{n_1 n_2, 1}$ , which are defined by

$$A_{n_h, 1} = \frac{1}{n_h(n_h - 1)} \sum_{i \neq j} (X'_{hi} X_{hj})^2, \quad C_{n_1 n_2, 1} = \frac{1}{n_1 n_2} \sum_{ij} (X'_{1i} X_{2j})^2.$$

Hence, we only need to study the asymptotic normality of  $Z_{n_1, n_2}$  which is defined by  $Z_{n_1, n_2} = A_{n_1, 1} + A_{n_2, 1} - 2C_{n_1 n_2, 1}$ .

In order to construct a martingale sequence, it is convenient to have new random variables  $Y_i$  which are defined as

$$\begin{aligned}
 Y_i &= X_{1i} & \text{for } i = 1, 2, \dots, n_1, \\
 Y_{n_1+j} &= X_{2j} & \text{for } j = 1, 2, \dots, n_2.
 \end{aligned}$$

To construct a martingale difference, we let  $\mathcal{F}_0 = \{\emptyset, \Omega\}$ ,  $\mathcal{F}_k = \sigma\{Y_1, \dots, Y_k\}$  with  $k = 1, 2, \dots, n_1 + n_2$ . And let  $E_k(\cdot)$  denote the conditional expectation given  $\mathcal{F}_k$ . Define  $D_{n, k} = (E_k - E_{k-1})Z_{n_1, n_2}$  and it is easy to see that  $Z_{n_1, n_2} - E(Z_{n_1, n_2}) = \sum_{k=1}^{n_1+n_2} D_{n, k}$ .

LEMMA 1. *For any  $n$ ,  $\{D_{n, k}, 1 \leq k \leq n\}$  is a martingale difference sequence with respect to the  $\sigma$ -fields  $\{\mathcal{F}_k, 1 \leq k \leq n\}$ .*

PROOF. First of all, it is straightforward to show that  $ED_{n, k} = 0$ . Next, by denoting  $S_{n, m} = \sum_{k=1}^m D_{n, k} = E_m Z_{n_1, n_2} - E Z_{n_1, n_2}$ , we have  $S_{n, q} = S_{n, m} + (E_q Z_{n_1, n_2} - E_m Z_{n_1, n_2})$ . Then we can show that  $E(S_{n, q} | \mathcal{F}_m) = S_{n, m}$ . This completes the proof of Lemma 1.  $\square$



To apply martingale central limit theorem, we need Lemmas 2 and 3.

LEMMA 2. *Under Condition A2 and as  $\min\{n_1, n_2\} \rightarrow \infty$ ,*

$$\frac{\sum_{k=1}^{n_1+n_2} \sigma_{n,k}^2}{\text{Var}(Z_{n_1, n_2})} \xrightarrow{p} 1,$$

where  $\sigma_{n,k}^2 = E_{k-1}(D_{n,k}^2)$ .

PROOF. To prove Lemma 2, first we can show  $E(\sum_{k=1}^{n_1+n_2} \sigma_{n,k}^2) = \text{Var}(Z_{n_1, n_2})$ . Then we will show that as  $\min\{n_1, n_2\} \rightarrow \infty$ ,  $\text{Var}(\sum_{k=1}^{n_1+n_2} \sigma_{n,k}^2) / \text{Var}^2(Z_{n_1, n_2}) \rightarrow 0$ . To this end, we decompose  $\sum_{k=1}^{n_1+n_2} \sigma_{n,k}^2$  into the sum of eight parts,

$$\sum_{k=1}^{n_1+n_2} \sigma_{n,k}^2 = R_1 + R_2 + R_3 + R_4 + R_5 + R_6 + R_7 + R_8,$$

where with  $Q_{1,k-1} = \sum_{i=1}^{k-1} (Y_i Y_i' - \Sigma_1)$  and  $Q_{2,n_1+l-1} = \sum_{i=1}^{l-1} (Y_{n_1+i} Y_{n_1+i}' - \Sigma_2)$ ,

$$\begin{aligned} R_1 &= \sum_{k=1}^{n_1} \frac{8}{n_1^2(n_1-1)^2} \text{tr}(Q_{1,k-1} \Sigma_1 Q_{1,k-1} \Sigma_1) \\ &\quad + \sum_{l=1}^{n_2} \frac{8}{n_2^2(n_2-1)^2} \text{tr}(Q_{2,n_1+l-1} \Sigma_2 Q_{2,n_1+l-1} \Sigma_2), \\ R_2 &= \sum_{k=1}^{n_1} \frac{16}{n_1^2(n_1-1)} \sum_{i=1}^{k-1} \{Y_i' (\Sigma_1^3 - \Sigma_1 \Sigma_2 \Sigma_1) Y_i\}, \\ R_3 &= \sum_{l=1}^{n_2} \frac{16}{n_2^2(n_2-1)} \left[ \text{tr}(Q_{2,n_1+l-1} \Sigma_2^3) - \text{tr} \left\{ Q_{2,n_1+l-1} \Sigma_2 \left( \frac{1}{n_1} \sum_{i=1}^{n_1} Y_i Y_i' \right) \Sigma_2 \right\} \right], \\ R_4 &= \frac{8}{n_1^2 n_2} \sum_{i,j}^{n_1} \text{tr}(Y_j Y_j' \Sigma_2 Y_i Y_i' \Sigma_2) - \frac{16}{n_1 n_2} \text{tr} \left\{ \Sigma_2^3 \left( \sum_{i=1}^{n_1} Y_i Y_i' \right) \right\}, \\ R_5 &= \sum_{k=1}^{n_1} \frac{4\Delta_1}{n_1^2(n_1-1)^2} \text{tr}(\Gamma_1' Q_{1,k-1} \Gamma_1 \circ \Gamma_1' Q_{1,k-1} \Gamma_1) \\ &\quad + \sum_{l=1}^{n_2} \frac{4\Delta_2}{n_2^2(n_2-1)^2} \text{tr}(\Gamma_2' Q_{2,n_1+l-1} \Gamma_2 \circ \Gamma_2' Q_{2,n_1+l-1} \Gamma_2), \\ R_6 &= \sum_{k=1}^{n_1} \frac{8\Delta_1}{n_1^2(n_1-1)} \text{tr}\{\Gamma_1' (\Sigma_1 - \Sigma_2) \Gamma_1 \circ \Gamma_1' Q_{1,k-1} \Gamma_1\}, \end{aligned}$$

$$R_7 = \sum_{l=1}^{n_2} \frac{8\Delta_2}{n_2^2(n_2-1)} \left[ \text{tr}(\Gamma'_2 Q_{2,n_1+l-1} \Gamma_2 \circ \Gamma'_2 \Sigma_2 \Gamma_2) \right. \\ \left. - \text{tr} \left\{ \Gamma'_2 Q_{2,n_1+l-1} \Gamma_2 \circ \Gamma'_2 \left( \frac{1}{n_1} \sum_{i=1}^{n_1} Y_i Y'_i \right) \Gamma_2 \right\} \right]$$

and

$$R_8 = \frac{4\Delta_2}{n_1^2 n_2} \sum_{i,j}^{n_1} \text{tr}(\Gamma'_2 Y_i Y'_i \Gamma_2 \circ \Gamma'_2 Y_j Y'_j \Gamma_2) - \frac{8\Delta_2}{n_1 n_2} \sum_{i=1}^{n_1} \text{tr}(\Gamma'_2 \Sigma_2 \Gamma_2 \circ \Gamma'_2 Y_i Y'_i \Gamma_2).$$

Therefore, we need to show that  $\text{Var}(R_i) = o\{\text{Var}^2(Z_{n_1, n_2})\}$  for  $i = 1, \dots, 8$ . For  $R_1$ , there exists a constant  $K_1$  such that

$$\text{Var}(R_1) \leq K_1 \{n_1^{-4} \text{tr}^2(\Sigma_1^2) \text{tr}(\Sigma_1^4) + n_2^{-4} \text{tr}^2(\Sigma_2^2) \text{tr}(\Sigma_2^4)\}.$$

Then, applying  $\text{Var}^2(Z_{n_1, n_2}) \geq \frac{16}{n_1^4} \text{tr}^4(\Sigma_1^2) + \frac{16}{n_2^4} \text{tr}^4(\Sigma_2^2)$  from (2.5), we know

$$\frac{\text{Var}(R_1)}{\text{Var}^2(Z_{n_1, n_2})} \leq \frac{K_1}{16} \left\{ \frac{\text{tr}(\Sigma_1^4)}{\text{tr}^2(\Sigma_1^2)} + \frac{\text{tr}(\Sigma_2^4)}{\text{tr}^2(\Sigma_2^2)} \right\},$$

where  $\text{tr}(\Sigma_1^4)/\text{tr}^2(\Sigma_1^2) \rightarrow 0$  under Condition A2. Thus,  $\text{Var}(R_1) = o\{\text{Var}^2(Z_{n_1, n_2})\}$ .

By carrying out similar procedures we can show that the above is true for  $R_i$  with  $i = 1, \dots, 8$ . Hence we complete the proof of Lemma 2.  $\square$

LEMMA 3. Under Condition A2, as  $\min\{n_1, n_2\} \rightarrow \infty$

$$\frac{\sum_{k=1}^{n_1+n_2} \mathbb{E}(D_{n,k}^4)}{\text{Var}^2(Z_{n_1, n_2})} \rightarrow 0.$$

PROOF. For the case of  $1 \leq k \leq n_1$ , there exists a constant  $c$  such that

$$\sum_{k=1}^{n_1} \mathbb{E}(D_{n,k}^4) \leq c[n_1^{-3} \text{tr}^2\{(\Sigma_1^2 - \Sigma_1 \Sigma_2)^2\} + n_1^{-5} \text{tr}^4\{(\Sigma_1^2)\}].$$

Using the results  $\text{Var}^2(Z_{n_1, n_2}) \geq 64n_1^{-2} \text{tr}^2\{(\Sigma_1^2 - \Sigma_1 \Sigma_2)^2\}$  and  $\text{Var}^2(Z_{n_1, n_2}) \geq 16n_1^{-4} \text{tr}^4\{(\Sigma_1^2)\}$  from (2.5) and as  $n_1 \rightarrow \infty$ , we have

$$\frac{\sum_{k=1}^{n_1} \mathbb{E}(D_{n,k}^4)}{\text{Var}^2(Z_{n_1, n_2})} \leq \frac{c}{n_1} \rightarrow 0.$$

For the case of  $n_1 < k < n_1 + n_2$ , there exists a constant  $d$  such that

$$\sum_{k=n_1}^{n_1+n_2} \mathbb{E}(D_{n,k}^4) \leq \frac{d}{n_1^2 n_2^4} \{2 \text{tr}^4(\Sigma_1 \Sigma_2) + \text{tr}^2(\Sigma_1 \Sigma_2) \text{tr}^2(\Sigma_1^2)\}$$

$$\begin{aligned}
(6.6) \quad & + \frac{d}{n_1 n_2^4} [2 \operatorname{tr}^2(\Sigma_1 \Sigma_2) \operatorname{tr}\{(\Sigma_2^2 - \Sigma_2 \Sigma_1)^2\}] + \frac{d}{n_2^5} \operatorname{tr}^4\{(\Sigma_2^2)\} \\
& + \frac{d}{n_2^4} [2 \operatorname{tr}^2(\Sigma_2^2) \operatorname{tr}\{(\Sigma_2^2 - \Sigma_2 \Sigma_1)^2\} + 4 \operatorname{tr}^2(\Sigma_1 \Sigma_2) \operatorname{tr}^2(\Sigma_2^2)].
\end{aligned}$$

To evaluate the ratio of individual term in (6.6) to  $\operatorname{Var}^2(Z_{n_1, n_2})$ , respectively, we simply replace  $\operatorname{Var}^2(Z_{n_1, n_2})$  by corresponding terms in (2.5). Then under Condition A2 and as  $n_2 \rightarrow \infty$ ,  $\sum_{k=n_1+1}^{n_1+n_2} \operatorname{E}(D_{n,k}^4) / \operatorname{Var}^2(Z_{n_1, n_2}) \rightarrow 0$ . Therefore, we complete the proof of Lemma 3.  $\square$

With two sufficient conditions given in Lemmas 2 and 3, we conclude that

$$\frac{Z_{n_1, n_2} - \operatorname{E}(Z_{n_1, n_2})}{\operatorname{Var}(Z_{n_1, n_2})} \xrightarrow{d} N(0, 1).$$

If we let  $\varepsilon_{n_1, n_2} = A_{n_1, 2} + A_{n_1, 3} + A_{n_2, 2} + A_{n_2, 3} - 2C_{n_1 n_1, 2} - 2C_{n_1 n_1, 3} - 2C_{n_1 n_1, 4}$ , then  $T_{n_1, n_2} = Z_{n_1, n_2} + \varepsilon_{n_1, n_2}$ . From  $\operatorname{Var}(\varepsilon_{n_1, n_2}) = o(\sigma_{n_1, n_2}^2)$ ,

$$\operatorname{Var}\left(\frac{\varepsilon_{n_1, n_2}}{\sigma_{n_1, n_2}}\right) = \frac{\operatorname{Var}(\varepsilon_{n_1, n_2})}{\sigma_{n_1, n_2}^2} \rightarrow 0.$$

Moreover,  $\operatorname{E}(\varepsilon_{n_1, n_2}) = 0$ . Therefore,  $\varepsilon_{n_1, n_2} / \sigma_{n_1, n_2} \xrightarrow{p} 0$ . From Slutsky's Theorem, we complete the proof of Theorem 1.

**6.3. Proof of Theorem 2.** Recall that  $\operatorname{E}(A_{n_h}) = \operatorname{tr}(\Sigma_h^2)$  for  $h = 1$  or  $2$ . To show  $A_{n_h} / \operatorname{tr}(\Sigma_h^2) \xrightarrow{p} 1$ , it is sufficient to show that  $\operatorname{Var}\{A_{n_h} / \operatorname{tr}(\Sigma_h^2)\} \rightarrow 0$ .

From (6.2), we have

$$\begin{aligned}
& \operatorname{Var}\left\{\frac{A_{n_h}}{\operatorname{tr}(\Sigma_h^2)}\right\} \\
& \leq \frac{1}{\operatorname{tr}^2(\Sigma_h^2)} \left[ \frac{4}{n_h^2} \operatorname{tr}^2(\Sigma_h^2) + \frac{8 + 4\Delta_h}{n_h} \operatorname{tr}(\Sigma_h^4) + O\left\{\frac{1}{n_h^3} \operatorname{tr}^2(\Sigma_h^2) + \frac{1}{n_h^2} \operatorname{tr}(\Sigma_h^4)\right\} \right],
\end{aligned}$$

where  $\operatorname{tr}(\Sigma_h^4) / \operatorname{tr}^2(\Sigma_h^2) \rightarrow 0$  under Condition A2. Hence,  $A_{n_h} / \operatorname{tr}(\Sigma_h^2) \xrightarrow{p} 1$ .

Moreover, under  $H_{0a} : \Sigma_1 = \Sigma_2 = \Sigma$ ,  $A_{n_h} / \operatorname{tr}(\Sigma^2) \xrightarrow{p} 1$ . Then using the continuous mapping theorem, we have  $\hat{\sigma}_{0, n_1, n_2} / \sigma_{0, n_1, n_2} \xrightarrow{p} 1$ .

**6.4. Proof of Theorem 3.** The leading order terms in  $\operatorname{Var}(S_{n_1, n_2})$  are contributed by  $U_{n_h, 1}$  and  $W_{n_1 n_2, 1}$  which are defined by

$$\begin{aligned}
U_{n_h, 1} &= \frac{1}{n_h(n_h - 1)} \sum_{i \neq j} X_{hi}^{(1)'} X_{hj}^{(1)} X_{hj}^{(2)'} X_{hi}^{(2)}, \\
W_{n_1 n_2, 1} &= \frac{1}{n_1 n_2} \sum_{ij} X_{1i}^{(1)'} X_{2j}^{(1)} X_{2j}^{(2)'} X_{1i}^{(2)}.
\end{aligned}$$

From Slutsky's Theorem, we only need to study the asymptotic normality of  $H_{n_1, n_2}$  which is defined as  $H_{n_1, n_2} =: U_{n_1, 1} + U_{n_2, 1} - 2W_{n_1 n_2, 1}$ .

To implement martingale central limit theorem to  $H_{n_1, n_2}$ , we need a martingale sequence. To this end, we define random variables which are

$$\begin{aligned} Y_i^{(1)} &= X_{1i}^{(1)} \quad \text{and} \quad Y_i^{(2)} = X_{1i}^{(2)} \quad \text{for } i = 1, 2, \dots, n_1, \\ Y_{n_1+j}^{(1)} &= X_{2j}^{(1)} \quad \text{and} \quad Y_{n_1+j}^{(2)} = X_{2j}^{(2)} \quad \text{for } j = 1, 2, \dots, n_2. \end{aligned}$$

If we define  $C_{n,k} = (E_k - E_{k-1})H_{n_1, n_2}$ , where  $E_k(\cdot)$  denote the conditional expectation given  $\mathcal{F}_k = \sigma\{Y_1, \dots, Y_k\}$  with  $k = 1, 2, \dots, n_1 + n_2$ , we claim that  $\{C_{n,k}, 1 \leq k \leq n\}$  is a martingale difference sequence with respect to the  $\sigma$ -fields  $\{\mathcal{F}_k, 1 \leq k \leq n\}$  from Lemma 1. We need Lemmas 4 and 5 to implement the martingale central limit theorem.

LEMMA 4. *Under Conditions A2 and A4, as  $\min\{n_1, n_2\} \rightarrow \infty$ ,*

$$\frac{\sum_{k=1}^{n_1+n_2} \tau_{n,k}^2}{\text{Var}(H_{n_1, n_2})} \xrightarrow{p} 1,$$

where  $\tau_{n,k}^2 = E_{k-1}(C_{n,k}^2)$ .

PROOF. First, we can show that  $E(\sum_{k=1}^{n_1+n_2} \tau_{n,k}^2) = \text{Var}(H_{n_1, n_2})$ . Therefore, we only need to show  $\text{Var}(\sum_{k=1}^{n_1+n_2} \tau_{n,k}^2) = o\{\text{Var}^2(H_{n_1, n_2})\}$  to complete the proof of Lemma 4. To this end, we decompose  $\sum_{k=1}^{n_1+n_2} \tau_{n,k}^2$  into twelve parts,

$$\sum_{k=1}^{n_1+n_2} \sigma_{n,k}^2 = P_1 + P_2 + P_3 + P_4 + P_5 + P_6 + P_7 + P_8 + P_9 + P_{10} + P_{11} + P_{12},$$

where with

$$\begin{aligned} O_{1,k-1} &= \sum_{i=1}^{k-1} (Y_i^{(1)} Y_i^{(2)'} - \Sigma_{1,12}) \quad \text{and} \\ O_{2,n_1+l-1} &= \sum_{i=1}^{l-1} (Y_{n_1+i}^{(1)} Y_{n_1+i}^{(2)'} - \Sigma_{2,12}), \\ P_1 &= \sum_{k=1}^{n_1} \frac{4}{n_1^2(n_1-1)^2} \text{tr}(O_{1,k-1} \Sigma'_{1,12} O_{1,k-1} \Sigma'_{1,12}) \\ &\quad + \sum_{l=1}^{n_2} \frac{4}{n_2^2(n_2-1)^2} \text{tr}(O_{2,n_1+l-1} \Sigma'_{2,12} O_{2,n_1+l-1} \Sigma'_{2,12}), \end{aligned}$$

$$\begin{aligned}
P_2 &= \sum_{k=1}^{n_1} \frac{4}{n_1^2(n_1-1)^2} \text{tr}(O_{1,k-1} \Sigma_{1,22} O'_{1,k-1} \Sigma_{1,11}) \\
&\quad + \sum_{l=1}^{n_2} \frac{4}{n_2^2(n_2-1)^2} \text{tr}(O_{2,n_1+l-1} \Sigma_{2,22} O'_{2,n_1+l-1} \Sigma_{2,11}), \\
P_3 &= \sum_{k=1}^{n_1} \frac{8}{n_1^2(n_1-1)} \text{tr}\{O_{1,k-1} \Sigma'_{1,12} (\Sigma_{1,12} - \Sigma_{2,12}) \Sigma'_{1,12}\}, \\
P_4 &= \sum_{k=1}^{n_1} \frac{8}{n_1^2(n_1-1)} \text{tr}\{O_{1,k-1} \Sigma_{1,22} (\Sigma'_{1,12} - \Sigma'_{2,12}) \Sigma_{1,11}\}, \\
P_5 &= \sum_{l=1}^{n_2} \frac{8}{n_2^2(n_2-1)} \text{tr}\left\{O_{2,n_1+l-1} \Sigma'_{2,12} \left(\Sigma_{2,12} - \frac{1}{n_1} \sum_{i=1}^{n_1} Y_i^{(1)} Y_i^{(2)'}\right) \Sigma'_{2,12}\right\}, \\
P_6 &= \sum_{l=1}^{n_2} \frac{8}{n_2^2(n_2-1)} \text{tr}\left\{O_{2,n_1+l-1} \Sigma_{2,22} \left(\Sigma'_{2,12} - \frac{1}{n_1} \sum_{i=1}^{n_1} Y_i^{(2)} Y_i^{(1)'}\right) \Sigma_{2,11}\right\}, \\
P_7 &= \frac{4}{n_2} \text{tr}\left\{\left(\Sigma_{2,12} - \frac{1}{n_1} \sum_{i=1}^{n_1} Y_i^{(1)} Y_i^{(2)'}\right) \Sigma'_{2,12}\right. \\
&\quad \left. \times \left(\Sigma_{2,12} - \frac{1}{n_1} \sum_{i=1}^{n_1} Y_i^{(1)} Y_i^{(2)'}\right) \Sigma'_{2,12}\right\}, \\
P_8 &= \frac{4}{n_2} \text{tr}\left\{\left(\Sigma_{2,12} - \frac{1}{n_1} \sum_{i=1}^{n_1} Y_i^{(1)} Y_i^{(2)'}\right) \Sigma_{2,22}\right. \\
&\quad \left. \times \left(\Sigma'_{2,12} - \frac{1}{n_1} \sum_{i=1}^{n_1} Y_i^{(2)} Y_i^{(1)'}\right) \Sigma_{2,11}\right\}, \\
P_9 &= \sum_{k=1}^{n_1} \frac{4\Delta_1}{n_1^2(n_1-1)^2} \text{tr}(\Gamma_1^{(1)'} O_{1,k-1} \Gamma_1^{(2)} \circ \Gamma_1^{(1)'} O_{1,k-1} \Gamma_1^{(2)}) \\
&\quad + \sum_{l=1}^{n_2} \frac{4\Delta_2}{n_2^2(n_2-1)^2} \text{tr}(\Gamma_2^{(1)'} O_{2,n_1+l-1} \Gamma_2^{(2)} \circ \Gamma_2^{(1)'} O_{2,n_1+l-1} \Gamma_2^{(2)}), \\
P_{10} &= \sum_{k=1}^{n_1} \frac{8\Delta_1}{n_1^2(n_1-1)} \text{tr}\{\Gamma_1^{(1)'} (\Sigma_{1,12} - \Sigma_{2,12}) \Gamma_1^{(2)} \circ \Gamma_1^{(1)'} O_{1,k-1} \Gamma_1^{(2)}\}, \\
P_{11} &= \sum_{l=1}^{n_2} \frac{8\Delta_2}{n_2^2(n_2-1)}
\end{aligned}$$

$$\begin{aligned}
& \times \operatorname{tr} \left\{ \Gamma_2^{(1)'} \left( \Sigma_{2,12} - \sum_{i=1}^{n_1} \frac{Y_i^{(1)} Y_i^{(2)'}}{n_1} \right) \Gamma_2^{(2)} \circ \Gamma_2^{(1)'} O_{2,n_1+l-1} \Gamma_2^{(2)} \right\}, \\
P_{12} &= \frac{4\Delta_2}{n_2} \operatorname{tr} \left\{ \Gamma_2^{(1)'} \left( \Sigma_{2,12} - \sum_{i=1}^{n_1} \frac{Y_i^{(1)} Y_i^{(2)'}}{n_1} \right) \Gamma_2^{(2)} \right. \\
& \quad \left. \circ \Gamma_2^{(1)'} \left( \Sigma_{2,12} - \sum_{i=1}^{n_1} \frac{Y_i^{(1)} Y_i^{(2)'}}{n_1} \right) \Gamma_2^{(2)} \right\}.
\end{aligned}$$

For  $P_1$ , there exists a constant  $J_1$  such that

$$\begin{aligned}
\operatorname{Var}(P_1) &\leq \sum_{h=1}^2 \frac{J_1}{n_h^4} \{ \operatorname{tr}^2(\Sigma_{h,12} \Sigma'_{h,12}) \operatorname{tr}(\Sigma_{h,11} \Sigma_{h,12} \Sigma_{h,22} \Sigma'_{h,12}) \\
& \quad + \operatorname{tr}(\Sigma_{h,11}^2) \operatorname{tr}(\Sigma_{h,22}^2) \operatorname{tr}(\Sigma_{h,11} \Sigma_{h,12} \Sigma_{h,22} \Sigma'_{h,12}) \\
& \quad + \operatorname{tr}^2(\Sigma_{h,11} \Sigma_{h,12} \Sigma_{h,22} \Sigma'_{h,12}) \}.
\end{aligned}$$

Using  $\operatorname{Var}^2(H_{n_1, n_2}) \geq \frac{8}{n_h^4} \operatorname{tr}(\Sigma_{h,11}^2) \operatorname{tr}(\Sigma_{h,22}^2) \operatorname{tr}^2(\Sigma_{h,12} \Sigma'_{h,12})$  from (3.8),

$$\begin{aligned}
& \frac{(J_1/(n_h^4)) \operatorname{tr}^2(\Sigma_{h,12} \Sigma'_{h,12}) \operatorname{tr}(\Sigma_{h,11} \Sigma_{h,12} \Sigma_{h,22} \Sigma'_{h,12})}{\operatorname{Var}^2(H_{n_1, n_2})} \\
& \leq \frac{J_1 \operatorname{tr}(\Sigma_{h,11} \Sigma_{h,12} \Sigma_{h,22} \Sigma'_{h,12})}{8 \operatorname{tr}(\Sigma_{h,11}^2) \operatorname{tr}(\Sigma_{h,22}^2)},
\end{aligned}$$

which goes to zero under Condition A4 for  $h = 1$  or  $2$ .

Similarly, using  $\operatorname{Var}^2(H_{n_1, n_2}) \geq \frac{4}{n_h^4} \operatorname{tr}^2(\Sigma_{h,11}^2) \operatorname{tr}^2(\Sigma_{h,22}^2)$  from (3.8),

$$\frac{J_1}{n_h^4} \operatorname{tr}^2(\Sigma_{h,11} \Sigma_{h,12} \Sigma_{h,22} \Sigma'_{h,12}) / \operatorname{Var}^2(H_{n_1, n_2}) \rightarrow 0, \quad \text{and}$$

$$\frac{J_1}{n_h^4} \operatorname{tr}(\Sigma_{h,11}^2) \operatorname{tr}(\Sigma_{h,22}^2) \operatorname{tr}(\Sigma_{h,11} \Sigma_{h,12} \Sigma_{h,22} \Sigma'_{h,12}) / \operatorname{Var}^2(H_{n_1, n_2}) \rightarrow 0.$$

Hence,  $\operatorname{Var}(P_1) = o\{\operatorname{Var}^2(H_{n_1, n_2})\}$ . Similarly, we have  $\operatorname{Var}(P_i) = o\{\operatorname{Var}^2(H_{n_1, n_2})\}$  for  $i = 1, \dots, 12$ . Therefore, we complete the proof of Lemma 4.  $\square$

LEMMA 5. Under Conditions A2 and A4, as  $\min\{n_1, n_2\} \rightarrow \infty$

$$\frac{\sum_{k=1}^{n_1+n_2} \mathbb{E}(C_{n,k}^4)}{\operatorname{Var}^2(H_{n_1, n_2})} \rightarrow 0.$$

PROOF. For the case of  $1 \leq k \leq n_1$ , there exists a constant  $c$  such that

$$\begin{aligned} \sum_{k=1}^{n_1} \mathbb{E}(C_{n,k}^4) &\leq c[n_1^{-3} \text{tr}^2\{\Sigma_{1,11}(\Sigma_{1,12} - \Sigma_{2,12})\Sigma_{1,22}(\Sigma'_{1,12} - \Sigma'_{2,12})\} \\ &\quad + n_1^{-5} \text{tr}^2(\Sigma_{1,11}^2) \text{tr}^2(\Sigma_{1,22}^2)]. \end{aligned}$$

Applying  $\text{Var}^2(H_{n_1, n_2}) \geq 16n_1^{-2} \text{tr}^2\{\Sigma_{1,11}(\Sigma_{1,12} - \Sigma_{2,12})\Sigma_{1,22}(\Sigma'_{1,12} - \Sigma'_{2,12})\}$  and  $\text{Var}^2(H_{n_1, n_2}) \geq 4n_1^{-4} \text{tr}^2(\Sigma_{1,11}^2) \text{tr}^2(\Sigma_{1,22}^2)$  from (3.8) and as  $n_1 \rightarrow \infty$ ,

$$\frac{\sum_{k=1}^{n_1} \mathbb{E}(C_{n,k}^4)}{\text{Var}^2(H_{n_1, n_2})} \leq \frac{c}{n_1} \rightarrow 0.$$

For the case of  $n_1 < k \leq n_1 + n_2$ , we can find a constant  $d$  such that

$$\begin{aligned} &\sum_{k=n_1}^{n_1+n_2} \mathbb{E}(C_{n,k}^4) \\ &\leq \frac{d}{n_1^3 n_2^3} \text{tr}(\Sigma_{1,11} \Sigma_{2,11}) \text{tr}(\Sigma_{1,22} \Sigma_{2,22}) \text{tr}(\Sigma_{2,11}^2) \text{tr}(\Sigma_{2,22}^2) \\ &\quad + \frac{d}{n_2^3} \text{tr}^2\{(\Sigma_{2,11} \Sigma_{2,12} - \Sigma_{2,11} \Sigma_{1,12})(\Sigma_{2,22} \Sigma'_{2,12} - \Sigma_{2,22} \Sigma'_{1,12})\} \\ (6.7) \quad &\quad + \frac{d}{n_1 n_2^3} \text{tr}(\Sigma_{1,11} \Sigma_{2,11}) \text{tr}(\Sigma_{1,22} \Sigma_{2,22}) \\ &\quad \times \text{tr}\{\Sigma_{2,11}(\Sigma_{2,12} - \Sigma_{1,12})\Sigma_{2,22}(\Sigma'_{2,12} - \Sigma'_{1,12})\} \\ &\quad + \frac{d}{n_1^2 n_2^3} \text{tr}^2(\Sigma_{1,11} \Sigma_{2,11}) \text{tr}^2(\Sigma_{1,22} \Sigma_{2,22}) + \frac{d}{n_2^5} \text{tr}^2(\Sigma_{2,11}^2) \text{tr}^2(\Sigma_{2,22}^2). \end{aligned}$$

To evaluate the ratio of individual term in (6.7) to  $\text{Var}^2(H_{n_1, n_2})$ , respectively, we simply replace  $\text{Var}^2(H_{n_1, n_2})$  by corresponding terms in (3.8). Then we can show that  $\sum_{k=n_1+1}^{n_1+n_2} \mathbb{E}(C_{n,k}^4) / \text{Var}^2(H_{n_1, n_2}) \rightarrow 0$ . Therefore, we complete the proof of Lemma 5.  $\square$

With two sufficient conditions given in Lemma 4 and 5, we know that

$$\frac{H_{n_1, n_2} - \mathbb{E}(H_{n_1, n_2})}{\text{Var}(H_{n_1, n_2})} \xrightarrow{d} N(0, 1).$$

If we let  $\varepsilon_{n_1, n_2} = U_{n_1, 2} + U_{n_1, 3} + U_{n_2, 2} + U_{n_2, 3} - 2W_{n_1 n_1, 2} - 2W_{n_1 n_1, 3} - 2W_{n_1 n_1, 4}$ , then  $S_{n_1, n_2} = H_{n_1, n_2} + \varepsilon_{n_1, n_2}$ . From  $\text{Var}(\varepsilon_{n_1, n_2}) = o(\sigma_{n_1, n_2}^2)$ ,

$$\text{Var}\left(\frac{\varepsilon_{n_1, n_2}}{\sigma_{n_1, n_2}}\right) = \frac{\text{Var}(\varepsilon_{n_1, n_2})}{\sigma_{n_1, n_2}^2} \rightarrow 0.$$

Moreover, we know  $\mathbb{E}(\varepsilon_{n_1, n_2}) = 0$ . Therefore,  $\varepsilon_{n_1, n_2} / \sigma_{n_1, n_2} \xrightarrow{p} 0$ . From Slutsky's Theorem, we complete the proof of Theorem 3.



6.5. *Proof of Theorem 4.* Applying the trace inequality, we know that  $\text{tr}^2(\Sigma_{h,12}\Sigma'_{h,12}) \leq \text{tr}(\Sigma_{h,11}^2)\text{tr}(\Sigma_{h,22}^2)$ . Therefore, to prove Theorem 4, we first consider the case where  $\text{tr}^2(\Sigma_{h,12}\Sigma'_{h,12}) = O\{\text{tr}(\Sigma_{h,11}^2)\text{tr}(\Sigma_{h,22}^2)\}$ . From Theorem 2, we can show that  $A_{n_h}^{(1)}/\text{tr}(\Sigma_{h,11}^2) \xrightarrow{p} 1$  and  $A_{n_h}^{(2)}/\text{tr}(\Sigma_{h,22}^2) \xrightarrow{p} 1$ . Moreover, from (6.3), there exists a constant  $d_1$  such that

$$\text{Var}\{C_{n_1n_2}^{(i)}/\text{tr}(\Sigma_{1,ii}\Sigma_{2,ii})\} \leq d_1 \left( \frac{1}{n_1} + \frac{1}{n_2} \right) \rightarrow 0,$$

which with  $E(C_{n_1n_2}^{(i)}) = \text{tr}(\Sigma_{1,ii}\Sigma_{2,ii})$  implies that  $C_{n_1n_2}^{(i)}/\text{tr}(\Sigma_{1,ii}\Sigma_{2,ii}) \xrightarrow{p} 1$ . Similarly, using  $\text{tr}^2(\Sigma_{h,12}\Sigma'_{h,12}) = O\{\text{tr}(\Sigma_{h,11}^2)\text{tr}(\Sigma_{h,22}^2)\}$ , we can find a constant  $d_2$  such that

$$\begin{aligned} & \text{Var}\{U_{n_h}/\text{tr}(\Sigma_{h,12}\Sigma'_{h,12})\} \\ & \leq \frac{d_2}{n_h} \{1 + \text{tr}(\Sigma_{h,11}^2)\text{tr}(\Sigma_{h,22}^2)/\text{tr}^2(\Sigma_{h,12}\Sigma'_{h,12})\} \\ & \rightarrow 0, \end{aligned}$$

which together with  $E(U_{n_h}) = \text{tr}(\Sigma_{h,12}\Sigma'_{h,12})$  shows that  $U_{n_h}/\text{tr}(\Sigma_{h,12}\Sigma'_{h,12}) \xrightarrow{p} 1$  for  $h = 1$  or  $2$ . Hence, if we define

$$\begin{aligned} \omega_{0,n_1,n_2,1}^2 &= 2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)^2 \text{tr}^2(\Sigma_{12}\Sigma'_{12}) \quad \text{and} \\ \omega_{0,n_1,n_2,2}^2 &= 2 \sum_{i=1}^2 \frac{1}{n_i^2} \text{tr}(\Sigma_{i,11}^2)\text{tr}(\Sigma_{i,22}^2) + \frac{4}{n_1n_2} \text{tr}(\Sigma_{1,11}\Sigma_{2,11})\text{tr}(\Sigma_{1,22}\Sigma_{2,22}), \end{aligned}$$

then under  $H_{0b}: \Sigma_{1,12} = \Sigma_{2,12} = \Sigma_{12}$  and from the mapping theorem,

$$\begin{aligned} \frac{\hat{\omega}_{0,n_1,n_2}^2}{\omega_{0,n_1,n_2}^2} &= \frac{\omega_{0,n_1,n_2,1}^2}{\omega_{0,n_1,n_2}^2} \frac{2(U_{n_1}/n_1 + U_{n_2}/n_2)^2}{\omega_{0,n_1,n_2,1}^2} \\ (6.8) \quad &+ \frac{\omega_{0,n_1,n_2,2}^2}{\omega_{0,n_1,n_2}^2} \frac{\sum_{i=1}^2 \{(2/n_i^2)A_{n_i}^{(1)}A_{n_i}^{(2)}\} + (4/(n_1n_2))C_{n_1n_2}^{(1)}C_{n_1n_2}^{(2)}}{\omega_{0,n_1,n_2,2}^2} \\ &\xrightarrow{p} 1. \end{aligned}$$

Next, we consider  $\text{tr}^2(\Sigma_{h,12}\Sigma'_{h,12}) = o\{\text{tr}(\Sigma_{h,11}^2)\text{tr}(\Sigma_{h,22}^2)\}$ . If we define

$$\begin{aligned} \hat{\omega}_{0,n_1,n_2,1}^2 &= 2 \left( \frac{U_{n_1}}{n_2} + \frac{U_{n_2}}{n_1} \right)^2 \quad \text{and} \\ \hat{\omega}_{0,n_1,n_2,2}^2 &= \sum_{i=1}^2 \left\{ \frac{2}{n_i} A_{n_i}^{(1)} A_{n_i}^{(2)} \right\} + \frac{4}{n_1n_2} C_{n_1n_2}^{(1)} C_{n_1n_2}^{(2)}, \end{aligned}$$

then, for a given constant  $\varepsilon$ , we have

$$P\left(\left|\frac{\hat{\omega}_{0,n_1,n_2}^2}{\omega_{0,n_1,n_2}^2} - 1\right| > \varepsilon\right) \leq P\left(\frac{\hat{\omega}_{0,n_1,n_2,1}^2}{\omega_{0,n_1,n_2}^2} > \varepsilon/2\right) + P\left(\left|\frac{\hat{\omega}_{0,n_1,n_2,2}^2}{\omega_{0,n_1,n_2}^2} - 1\right| > \varepsilon/2\right).$$

Thus, we only need to show  $\hat{\omega}_{0,n_1,n_2,1}^2/\omega_{0,n_1,n_2}^2 \xrightarrow{P} 0$  and  $\hat{\omega}_{0,n_1,n_2,2}^2/\omega_{0,n_1,n_2}^2 \xrightarrow{P} 1$ , respectively. First of all, we know  $\hat{\omega}_{0,n_1,n_2,2}^2/\omega_{0,n_1,n_2}^2 \xrightarrow{P} 1$  from (6.8). Second, there exists a constant  $d_3$  such that

$$P\left(\frac{\hat{\omega}_{0,n_1,n_2,1}^2}{\omega_{0,n_1,n_2}^2} > \frac{\varepsilon}{2}\right) \leq d_3 \left[ \frac{\sum_{i=1}^2 \text{tr}^2(\Sigma_{i,12} \Sigma'_{i,12})}{\sum_{i=1}^2 \text{tr}(\Sigma_{i,11}^2) \text{tr}(\Sigma_{i,22}^2)} + \sum_{i=1}^2 \left\{ \frac{1}{n_i} + \frac{\text{tr}^2(\Sigma_{i,12} \Sigma'_{i,12})}{n_1 \text{tr}(\Sigma_{i,11}^2) \text{tr}(\Sigma_{i,22}^2)} \right\} \right],$$

which converges to zero under  $\text{tr}^2(\Sigma_{i,12} \Sigma'_{i,12}) = o\{\text{tr}(\Sigma_{i,11}^2) \text{tr}(\Sigma_{i,22}^2)\}$ . Therefore, we have  $\hat{\omega}_{0,n_1,n_2}^2/\omega_{0,n_1,n_2}^2 \xrightarrow{P} 1$ , as claimed by Theorem 4.

**Acknowledgment.** We thank three referees for constructive comments and suggestions which have improved the presentation of the paper.

## REFERENCES

- ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. Wiley, Hoboken, NJ. [MR1990662](#)
- BAI, Z. D. (1993). Convergence rate of expected spectral distributions of large random matrices. II. Sample covariance matrices. *Ann. Probab.* **21** 649–672. [MR1217560](#)
- BAI, Z. and SARANADASA, H. (1996). Effect of high dimension: By an example of a two sample problem. *Statist. Sinica* **6** 311–329. [MR1399305](#)
- BAI, Z. and SILVERSTEIN, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. Springer, New York. [MR2567175](#)
- BAI, Z. D. and YIN, Y. Q. (1993). Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix. *Ann. Probab.* **21** 1275–1294. [MR1235416](#)
- BAI, Z., JIANG, D., YAO, J.-F. and ZHENG, S. (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *Ann. Statist.* **37** 3822–3840. [MR2572444](#)
- BARRY, W. T., NOBEL, A. B. and WRIGHT, F. A. (2005). Significance analysis of functional categories in gene expression studies: A structured permutation approach. *Bioinformatics* **21** 1943–1949.
- BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57** 289–300. [MR1325392](#)
- BICKEL, P. J. and LEVINA, E. (2008a). Regularized estimation of large covariance matrices. *Ann. Statist.* **36** 199–227. [MR2387969](#)
- BICKEL, P. J. and LEVINA, E. (2008b). Covariance regularization by thresholding. *Ann. Statist.* **36** 2577–2604. [MR2485008](#)
- CAI, T. T. and JIANG, T. (2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *Ann. Statist.* **39** 1496–1525. [MR2850210](#)

- CAI, T., LIU, W. D. and XIA, Y. (2011). Two-sample covariance matrix testing and support recovery. Technical report, Dept. Statistics, Univ. Pennsylvania, Philadelphia, PA.
- CHEN, S. X. and QIN, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.* **38** 808–835. [MR2604697](#)
- CHEN, S. X., ZHANG, L.-X. and ZHONG, P.-S. (2010). Tests for high-dimensional covariance matrices. *J. Amer. Statist. Assoc.* **105** 810–819. [MR2724863](#)
- CHIARETTI, S., LI, X. C., GENTLEMAN, R., VITALE, A., VIGNETTI, M., MANDELLI, F., RITZ, J. and FOA, R. (2004). Gene expression profile of adult T-cell acute lymphocytic leukemia identifies distinct subsets of patients with different response to therapy and survival. *Blood* **103** 2771–2778.
- DONOHU, D. and JIN, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32** 962–994. [MR2065195](#)
- DUDOIT, S., KELES, S. and VAN DER LAAN, M. J. (2008). Multiple tests of association with biological annotation metadata. In *Probability and Statistics: Essays in Honor of David A. Freedman. Inst. Math. Stat. Collect.* **2** 153–218. IMS, Beachwood, OH. [MR2459952](#)
- DYKSTRA, R. L. (1970). Establishing the positive definiteness of the sample covariance matrix. *Ann. Math. Statist.* **41** 2153–2154.
- EFRON, B. and TIBSHIRANI, R. (2007). On testing the significance of sets of genes. *Ann. Appl. Stat.* **1** 107–129. [MR2393843](#)
- EL KAROUI, N. (2007). Tracy–Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *Ann. Probab.* **35** 663–714. [MR2308592](#)
- FAN, J., FAN, Y. and LV, J. (2008). High dimensional covariance matrix estimation using a factor model. *J. Econometrics* **147** 186–197. [MR2472991](#)
- FAN, J., HALL, P. and YAO, Q. (2007). To how many simultaneous hypothesis tests can normal, Student's  $t$  or bootstrap calibration be applied? *J. Amer. Statist. Assoc.* **102** 1282–1288. [MR2372536](#)
- FAN, J., PENG, H. and HUANG, T. (2005). Semilinear high-dimensional model for normalization of microarray data: A theoretical analysis and partial consistency. *J. Amer. Statist. Assoc.* **100** 781–813. [MR2201010](#)
- GLASSER, G. J. (1961). An unbiased estimator for powers of the arithmetic mean. *J. Roy. Statist. Soc. Ser. B* **23** 154–159. [MR0123388](#)
- GLASSER, G. J. (1962). Estimators for the product of arithmetic means. *J. Roy. Statist. Soc. Ser. B* **24** 180–184. [MR0137209](#)
- HALL, P. and JIN, J. (2008). Properties of higher criticism under strong dependence. *Ann. Statist.* **36** 381–402. [MR2387976](#)
- HUANG, J., WANG, D. and ZHANG, C.-H. (2005). A two-way semilinear model for normalization and analysis of cDNA microarray data. *J. Amer. Statist. Assoc.* **100** 814–829. [MR2201011](#)
- HUANG, J. Z., LIU, N., POURAHMADI, M. and LIU, L. (2006). Covariance matrix selection and estimation via penalised normal likelihood. *Biometrika* **93** 85–98. [MR2277742](#)
- JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. [MR1863961](#)
- JOHNSTONE, I. M. and LU, A. Y. (2009). On consistency and sparsity for principal components analysis in high dimensions. *J. Amer. Statist. Assoc.* **104** 682–693. [MR2751448](#)
- LAM, C. and YAO, Q. (2011). Factor modelling for high-dimensional time series: Inference for the number of factors. *Ann. Statist.* To appear.
- LAM, C., YAO, Q. and BATHIA, N. (2011). Estimation of latent factors for high-dimensional time series. *Biometrika* **98** 901–918. [MR2860332](#)

- LAN, W., LUO, R., TSAI, C., WANG, H. and YANG, Y. (2010). Testing the diagonality of a large covariance matrix in a regression setting. Technical report, Peking Univ., China.
- LEDOIT, O. and WOLF, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist.* **30** 1081–1102. [MR1926169](#)
- LEDOIT, O. and WOLF, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivariate Anal.* **88** 365–411. [MR2026339](#)
- NETTLETON, D., RECKNOR, J. and REECY, J. M. (2008). Identification of differentially expressed gene categories in microarray studies using nonparametric multivariate analysis. *Bioinformatics* **24** 192–201.
- NEWTON, M. A., QUINTANA, F. A., DEN BOON, J. A., SENGUPTA, S. and AHLQUIST, P. (2007). Random-set methods identify distinct aspects of the enrichment signal in gene-set analysis. *Ann. Appl. Stat.* **1** 85–106. [MR2393842](#)
- ROTHMAN, A. J., LEVINA, E. and ZHU, J. (2010). A new approach to Cholesky-based covariance regularization in high dimensions. *Biometrika* **97** 539–550. [MR2672482](#)
- SCHOTT, J. R. (2007). A test for the equality of covariance matrices when the dimension is large relative to the sample sizes. *Comput. Statist. Data Anal.* **51** 6535–6542. [MR2408613](#)
- SHEDDEN, K. and TAYLOR, J. (2004). Differential correlation detects complex associations between gene expression and clinical outcomes in lung adenocarcinomas. In *Methods of Microarray Data Analysis IV* (J. S. SHOEMAKER and S. M. LIN, eds.) 121–131. Springer, New York.
- TRACY, C. A. and WIDOM, H. (1996). On orthogonal and symplectic matrix ensembles. *Comm. Math. Phys.* **177** 727–754. [MR1385083](#)
- VAN DER LAAN, M. J. and BRYAN, J. (2001). Gene expression analysis with the parametric bootstrap. *Biostatistics* **2** 445–461.
- WU, W. B. and POURAHMADI, M. (2003). Nonparametric estimation of large covariance matrices of longitudinal data. *Biometrika* **90** 831–844. [MR2024760](#)
- ZHANG, C.-H. and HUANG, J. (2008). The sparsity and bias of the LASSO selection in high-dimensional linear regression. *Ann. Statist.* **36** 1567–1594. [MR2435448](#)

DEPARTMENT OF STATISTICS  
IOWA STATE UNIVERSITY  
AMES, IOWA 50011-1210  
USA  
E-MAIL: [junli@iastate.edu](mailto:junli@iastate.edu)

GUANGHUA SCHOOL OF MANAGEMENT  
AND CENTER FOR STATISTICAL SCIENCE  
PEKING UNIVERSITY  
BEIJING 100871  
CHINA  
AND  
DEPARTMENT OF STATISTICS  
IOWA STATE UNIVERSITY  
AMES, IOWA 50011-1210  
USA  
E-MAIL: [csx@gsm.pku.edu.cn](mailto:csx@gsm.pku.edu.cn)